LOAN DOCUMENT

PHOTOGRAPH THIS SHEET

DTIC ACCESSION NUMBER

LEVEL

0

INVENTORY

Requirements Analysis for . . .

DOCUMENT IDENTIFICATION

MAR 89

DISTRIBUTION STATEMENT A
Approved for Public Release
Distribution Unlimited

DISTRIBUTION STATEMENT

H A N D L E W I T H C A R E

ACCESSION FOR
NTIS GRA&I ☑
DTIC TRAC ☐
UNANNOUNCED ☐
JUSTIFICATION

BY
DISTRIBUTION/
AVAILABILITY CODES

DISTRIBUTION	AVAILABILITY AND/OR SPECIAL	
A-1		

DISTRIBUTION STAMP

DATE ACCESSIONED

Reproduced From Best Available Copy

DATE RETURNED

20000203 115

DATE RECEIVED IN DTIC

REGISTERED OR CERTIFIED NUMBER

PHOTOGRAPH THIS SHEET AND RETURN TO DTIC-FDAC

DTIC FORM JUN 90 70A

DOCUMENT PROCESSING SHEET

PREVIOUS EDITIONS MAY BE USED UNTIL STOCK IS EXHAUSTED.

LOAN DOCUMENT

FOR FURTHER INFORMATION CONCERNING DISTRIBUTION CALL (703) 767-8040

PLEASE CHECK THE APPROPRIATE BLOCK BELOW:

AQ# ______

☐ ______ copies are being forwarded. Indicate whether Statement A, B, C, D, E, F, or X applies.

☑ DISTRIBUTION STATEMENT A:
APPROVED FOR PUBLIC RELEASE: DISTRIBUTION IS UNLIMITED

☐ DISTRIBUTION STATEMENT B:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES ONLY; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT C:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND THEIR CONTRACTORS; (Indicate Reason and Date). OTHER REQUESTS FOR THIS DOCUMENT SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT D:
DISTRIBUTION AUTHORIZED TO DoD AND U.S. DoD CONTRACTORS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT E:
DISTRIBUTION AUTHORIZED TO DoD COMPONENTS ONLY; (Indicate Reason and Date). OTHER REQUESTS SHALL BE REFERRED TO (Indicate Controlling DoD Office).

☐ DISTRIBUTION STATEMENT F:
FURTHER DISSEMINATION ONLY AS DIRECTED BY (Indicate Controlling DoD Office and Date) or HIGHER DoD AUTHORITY.

☐ DISTRIBUTION STATEMENT X:
DISTRIBUTION AUTHORIZED TO U.S. GOVERNMENT AGENCIES AND PRIVATE INDIVIDUALS OR ENTERPRISES ELIGIBLE TO OBTAIN EXPORT-CONTROLLED TECHNICAL DATA IN ACCORDANCE WITH DoD DIRECTIVE 5230.25. WITHHOLDING OF UNCLASSIFIED TECHNICAL DATA FROM PUBLIC DISCLOSURE, 6 Nov 1984 (indicate date of determination). CONTROLLING DoD OFFICE IS (Indicate Controlling DoD Office).

☐ This document was previously forwarded to DTIC on ______ (date) and the AD number is ______

☐ In accordance with provisions of DoD instructions, the document requested is not supplied because:

☐ It will be published at a later date. (Enter approximate date, if known).

☐ Other. (Give Reason)

DoD Directive 5230.24, "Distribution Statements on Technical Documents," 18 Mar 87, contains seven distribution statements, as described briefly above. Technical Documents must be assigned distribution statements.

Cynthia Gleisberg
Print or Type Name

334-?255-2924
Telephone Number

Cynthia Gleisberg
Authorized Signature/Date

PNL-6819
Limited Distribution

Requirements Analysis for the Army Safety Management Information System (ASMIS) Final Report

J. S. Littlefield
A. L. Corrigan

March 1989

Prepared for the
U.S. Army Safety Center
Ft. Rucker, Alabama
under a Related Services Agreement
with the U.S. Department of Energy
Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Operated for the U.S. Department of Energy
by Battelle Memorial Institute

Battelle

PNL-6819

CON 89-03

DISCLAIMER

PACIFIC NORTHWEST LABORATORY
operated by
BATTELLE MEMORIAL INSTITUTE
for the
UNITED STATES DEPARTMENT OF ENERGY
under Contract DE-AC06-76RLO 1830

PNL-6819

REQUIREMENTS ANALYSIS
FOR THE ARMY SAFETY MANAGEMENT
INFORMATION SYSTEM (ASMIS)
FINAL REPORT

J. S. Littlefield
A. L. Corrigan

March 1989

Prepared for the
U. S. Army Safety Center
Ft. Rucker, Alabama
under a Related Services Agreement
with the U.S. Department of Energy
Contract DE-AC06-76RLO 1830

Pacific Northwest Laboratory
Richland, Washington 99352

SUMMARY

This study for the U.S. Army Safety Center was undertaken to determine the requirements for an enhanced ASMIS system that could provide effective support to the Army world-wide safety community. Since the last major redesign of the ASMIS system in 1981, use of the system, particularly ad hoc requests, has been increasing and is expected to continue to increase. The goal of the redesign of the ASMIS system is to maximize the productivity of the users of the system by more effectively utilizing the available computing facilities.

This study involved examination and documentation of the existing ASMIS system, identification of problems areas within the current implementation, development of functional requirements for an enhanced system, proposal of multiple implementations, selection of the best implementation, evaluation of potential difficulties associated with the chosen implementation, and development of recommendations for the implementation of the enhanced ASMIS system.

The major conclusion of this study is that the ASMIS system should be reimplemented using DATACOM/DB, a relational database management system. DATACOM/DB is immediately available through a site license and is free to all Army installations. Conversion to a database management system would produce a significantly different style of operation for both the developers and users of the ASMIS system. Details of this proposed change and the associated implications are explained below.

Like the existing system, the reimplemented ASMIS system would have four components: an ad hoc query facility, a routine reporting facility, a data entry facility and a facility to precompute often-used statistics during times of low system utilization. For each of these facilities, the database management system (DBMS) would be used to provide data organization and storage, and to select the database cases.

A critical feature of the DBMS-based implementation would be the elimination of repetitious work in the ad hoc query facility. This could be accomplished by changing the paradigm for this facility from the current

repetitively used "get the database cases and create a single report or summarization" to a single invocation of "select a subset of database cases," followed by multiple invocations of "generate a report or summarization." Further reduction in work would come from allowing users to store sets of selected data on disk for later reuse. Additional reduction in the IBM 4381 workload could be accomplished by using personal computers as analysis workstations, thus offloading much of the analysis work. To facilitate this structure, specific recommendations are:

- A database management system should be used to select the database cases specified by the user's selection criteria.
- A statistical analysis package should be used to provide the summarizations. SAS, which is available on the IBM 4381, could be used.
- A reporting package should be used to provide columnar listings and caseprints. SAS or a package associated with DATACOM/DB could be used.
- For frequent users, personal computers could serve as analysis workstations, offloading much of the analysis and reporting work. Statistical analysis and reporting tools which provide the same functionality as the tools on the mainframe must be available. SAS PC could provide the necessary summarizations. SAS PC or PC DATACOM could provide the reporting capabilities. In addition, many other tools, such as LOTUS 1-2-3, DBASE, etc., could be used by personnel familiar with them.
- For infrequent users and those without personal computers, the IBM 4381 would provide the analysis and reporting functions as well as disk storage for data subsets.

The routine reporting facility and the facility to precompute statistics would remain relatively unchanged, except that they would run against the DBMS instead of accessing a series of individual data files. More options for coding routine reports would exist; fourth generation tools accompanying DATACOM/DB, the ad hoc query facility and reporting and summarization tools, and COBOL could be used. The ad hoc queries should be monitored and, as necessary, the precomputed statistics should be expanded to include additional often-used information.

The data entry facility would be changed. Because users can save selected subsets of data for multiple analyses, the requirement that the database not change during the day no longer exists. On-line, real-time updating of the database is recommended. This would reduce the complexity of the data entry facility and enable the viewing of all data as it is updated.

To allow the DBMS to access to all data, both current and historical, all ASMIS data would be stored on-line. To provide adequate storage for existing data and user created data subsets, and to provide space for short-term growth, purchase of another 2.5 gigabyte disk drive is recommended.

Making this change to the structure of the ASMIS system would necessitate changes for both the USASC programming staff and the ASMIS users. Effective implementation and use of the DBMS-based ASMIS system will require reorientation and training for the USASC programming staff. Four positions within DOIM should be identified, filled and the people trained: a database administrator (DBA), a backup for the DBA, a statistical analysis expert and a personal computer expert. For the users, productive, efficient use of the system will require training. Users must understand the organization of the data, the content of the data, and the tools available for manipulating that data. Training is particularly important for the USASC staff; approximately 70 percent of the queries come from within the USASC.

Other major recommendations for the Army Safety Center are:

- System security should be improved. An access control system should be added to the IBM 4381 to protect against intentional/unintentional destruction/corruption of data and programs. Each user should exist in a captive environment, having access to only the ASMIS system and his own data files. An on-line password facility should be implemented to allow a user to change his password immediately, independent of USASC personnel.

- Data communications should be improved. Dial-in communications should be expanded to support 2400 baud. Protocol checked transmissions to and from the IBM 4381 should be provided. Use of SIMPC on personal computers in conjunction with SIM 3278/VTAM on the IBM 4831 provides protocol checked transmission. For users without personal computers, MNP modems at USASC and the user site, would provide protocol checked transmissions.

CONTENTS

SUMMARY iii

1.0 OVERVIEW 1.1

1.1 BACKGROUND 1.2

1.2 PURPOSE OF THIS STUDY 1.5

1.3 SCOPE OF THIS DOCUMENT 1.5

2.0 SUMMARY AND RECOMMENDATIONS 2.1

2.1 SUMMARY 2.1

2.1.1 Results of Investigations 2.2

2.1.2 A Scenario for the Enhanced ASMIS System 2.5

2.2. RECOMMENDATIONS 2.6

2.2.1 Computer Software 2.9

2.2.2 Disk Storage 2.10

2.2.3 Organizational Structure 2.10

2.2.4 Training Requirements 2.12

2.2.5 Structure of the New ASMIS Implementation 2.13

2.2.6 Other Recommendations 2.17

3.0 THE EXISTING SYSTEM 3.1

3.1 DATA FLOW 3.3

3.1.1 Interfaces External to the Safety Center 3.3

3.1.2 Interfaces Internal to the Safety Center 3.5

3.2 DATABASES 3.8

3.2.1 Ground Accident Database 3.9

3.2.2 Aviation Accident Database 3.13

3.2.3 FECA Accident Database 3.17

CONTENTS (Cont'd)

3.2.4 Drug and Alcohol Database 3.19

3.2.5 Files Common to Multiple Databases 3.22

3.3 ROUTINE PROCESSING 3.23

3.3.1 Data Entry 3.24

3.3.2 Reporting 3.25

3.3.3 Statistics Generation 3.25

3.3.4 Safety Goal Generation 3.28

3.4 ACCESS METHODS USED TO RESPOND TO USER REQUESTS 3.28

3.4.1 ARPS 3.28

3.4.2 Non-ARPS 3.30

3.5 DATA ANALYSIS FACILITIES AVAILABLE TO USERS 3.30

3.6 USER DOCUMENTATION 3.30

3.7 DATA SECURITY MECHANISMS 3.30

3.7.1 Backup and Archiving of Data 3.30

3.7.2 File Protection 3.31

3.7.3 ARPS Security Features 3.31

3.8 COMPUTER SYSTEM AND SOFTWARE 3.31

3.8.1 Computer System 3.31

3.8.2 Computer Software 3.34

4.0 FUNCTIONAL REQUIREMENTS 4.1

4.1 COLLECTION AND COMPUTERIZATION OF THE DATA 4.1

4.1.1 Data Acquisition and Preparation 4.1

4.1.2 Data Entry 4.3

CONTENTS (Cont'd)

4.2 THE DATABASE SYSTEM 4.4

4.2.1 Relationship Between the Database Redesign and the Data Forms 4.4

4.2.2 Data Integration 4.5

4.2.3 Type of Database Access 4.6

4.2.4 Data Selection Capability 4.7

4.2.5 Response Time 4.8

4.2.6 Database Flexibility 4.9

4.2.7 Maintenance of the Database 4.11

4.3. USER INTERFACE 4.12

4.3.1 Data Accessibility 4.12

4.3.2 Query Specification 4.13

4.3.3 Report Facility 4.14

4.3.4 Isolation of a Subset of Data 4.16

4.3.5 Alternative Computing Resources 4.16

4.3.6 Maintainable User Applications 4.17

4.3.7 User Documentation 4.19

4.4. COMPUTING ENVIRONMENT 4.20

4.4.1 Data Security 4.20

4.4.2 System Security 4.21

4.4.3 Data Backup 4.22

4.4.4 System Backup 4.23

4.4.5 Data Communications 4.24

5.0 POSSIBLE IMPLEMENTATIONS 5.1

CONTENTS (Cont'd)

5.1 EXISTING ASMIS SYSTEM 5.3

5.1.1 Features 5.3

5.1.2 Advantages 5.5

5.1.3 Disadvantage 5.5

5.2 CENTRALIZED SINGLE USER QUERY PROCESSOR 5.5

5.2.1 Features 5.6

5.2.2 Advantages 5.6

5.2.3 Disadvantages 5.8

5.3 CENTRALIZED MULTIPLE USER QUERY PROCESSOR 5.9

5.3.1 Features 5.11

5.3.2 Advantages 5.11

5.3.3 Disadvantages 5.12

5.4 DATABASE MANAGEMENT SYSTEM (DBMS) 5.13

5.4.1 Features 5.14

5.4.2 Advantages 5.14

5.4.3 Disadvantages 5.15

5.5 DBMS WITH MULTIPLE USERS QUERY PROCESSOR 5.16

5.5.1 Features 5.18

5.5.2 Advantages 5.18

5.5.3 Disadvantages 5.20

5.6 SUMMARY OF POSSIBLE IMPLEMENTATIONS 5.21

6.0 EVALUATIONS 6.1

6.1 PERFORMANCE ANALYSIS OF THE CURRENT SYSTEM 6.1

6.1.1 Summary and Conclusions 6.1

CONTENTS (Cont'd)

6.1.2 Procedure 6.2

6.1.3 Detailed Findings and Discussion 6.3

6.2 SELECTING A DATABASE MODEL 6.8

6.2.1 Class I Environment: Files 6.8

6.2.2 Class II Environment: Transaction-oriented Databases 6.8

6.2.3 Class III Environment: Information systems 6.9

6.2.4 The Current ARPS Data Storage Environment 6.9

6.3 DATABASE EVALUATION 6.10

6.3.1 DB2 . . . . 6.11

6.3.2 Oracle 6.12

6.3.3 Cullinet 6.12

6.4 DATACOM/DB VERSUS SUPRA 6.13

6.4.1 Software and Services Available 6.15

6.4.2 SQL 6.17

6.4.3 PC to Mainframe Support 6.17

6.4.4 Variable Length for Narrative Data 6.17

6.4.5 Performance 6.17

6.5 DATACOM/DB TRAINING 6.18

6.6 MVS/SP VERSUS MVS/XA 6.19

6.7 ARPS QUERIES 6.20

6.7.1 Ground Database 6.22

6.7.2 Aviation Database 6.36

6.7.3 FECA, Flying Hours and Exposure Databases 6.45

6.7.4 Summary of Performance Evaluation 6.46

CONTENTS (Cont'd)

6.8 DISK SPACE UTILIZATION 6.48

REFERENCES . . . . Ref. 1

APPENDIX A: GLOSSARY A.1

APPENDIX B: CURRENT FORMS B.1

APPENDIX C: FIELDS USED IN ARGUMENTS AND FOR DISPLAY C.1

TABLES

3.1 Data Files Used, Number of Current and Historical Records, and and Growth Rate for the Ground Accident Database 3.12

3.2 Auxiliary Files Associated with the Ground Accident Database . . 3.12

3.3 Typical Reports Generated From The Ground Accident Database . . 3.13

3.4 Data Files Used, Number of Current and Historical Records, and Growth Rate for the Aviation Accident Database 3.16

3.5 Auxiliary Files Associated with the Aviation Accident Database . 3.16

3.6 Typical Reports Generated From the Aviation Accident Database . 3.17

3.7 Data Files Used, Number of Current and Historical Records, and Growth Rate for the FECA Accident Database 3.17

3.8 Auxiliary Files Associated with the FECA Accident Database . . . 3.19

3.9 Typical Reports Generated from the FECA Accident Database . . . 3.19

3.10 Data Files Used, Number of Current and Historical Records, and Growth Rate for the Drug and Alcohol Database 3.22

3.11 Auxiliary files associated with the drug and alcohol database . 3.22

3.12 Typical Reports Generated from the Drug and Alcohol Database . . 3.22

3.13 Data Files Used, Number of Current and Historical Records, and Growth Rate for Information Associated with Multiple Databases 3.23

5.1 Appropriateness of Possible Implementations on Saturated and Non-Saturated Systems 5.21

5.2 Comparison of Each of the Five Implementations Versus the Functional Requirements Where Differences Exist 5.24

5.3 Comparison of the Five Implementations Versus Issues Outside the Scope of the Functional Requirements. 5.25

5.4 Summary of the Comparison of the Five Implementations Versus the Functional Requirements and Administrative Issues 5.25

6.1 Options for New Product Purchases Required by MVS/XA 6.20

6.2 MVS/SP Products Displaced by Upgrading to MVS/XA 6.20

FIGURES

2.0	Conceptual View of Proposed System	2.8
3.0	Legend for Other Figures	3.2
3.1	USASC External Interfaces	3.4
3.2	USASC Internal Interfaces	3.6
3.3	Ground Database	3.10
3.4	Ground Daily Data Entry	3.11
3.5	Aviation Database	3.14
3.6	Aviation Daily Data Entry	3.15
3.7	FECA Database	3.18
3.8	Drug and Alcohol Database	3.20
3.9	Drug and Alcohol Daily Data Entry	3.21
3.10	Generation of Statistics File for Use in ARPS	3.26
3.11	Creation of Safety Goal Numbers	3.33
3.12	USASC Computer Hardware Configuration	3.37
5.0	Legend for Other Figures	5.2
5.1	Existing ASMIS System	5.4
5.2	Centralized Query Processor	5.7
5.3	Multiple User Query Processor	5.10
5.4	Database Management System (DBMS)	5.17
5.5	DBMS With Multiple User Query Processor	5.19

1.0 OVERVIEW

The current implementation of ASMIS and the ASMIS Retrieval and Processing System (ARPS) were developed in 1981. Since that time, it has adapted to fit the continually changing environment provided by the Army accident prevention community. Some of these adaptations were easy to make because the direction of change was foreseen, but the environment also evolved in unforseen directions. Other changes were very hard, if not impossible to make. During this time, the Safety Center has continually applied its five-step process of accident prevention. Monitoring the use of countermeasures and evaluating actual performance against projected benefits has required an ever increasing amount of information about each accident. Because many programs use the data files, revising the formats is very difficult. Adding information within the confines of these formats has resulted in data file formats which are very complex. The ASMIS system has outgrown its data file design.

During the years that ASMIS has been in use, the cost of computing equipment and personnel have changed dramatically. In the early 1980s, computing resources were expensive. ASMIS was developed to maximize the strong points of the then-current hardware and minimize the stress on its weak points, specifically disk storage and the speed of data transfer from disk to CPU. Since 1981, the cost of computer equipment has dropped to approximately one-tenth of the original (witness the proliferation of the personal computer) while personnel costs have increased approximately forty percent. Due to the environment existing at the time of its design, ASMIS is a very personnel intensive system. Thus, its current development is being limited by personnel costs and the number of people available to work on it.

With the advent of inexpensive computing, many more individuals have gained personal experience with computing. This has generated an increased demand for access to information, including the Army accident records. This has resulted in increased demands on the ASMIS system.

In April 1988, the IBM 4341 computer was replaced with an IBM 4381. Since its installation, the effect of the faster CPU and increased memory has been positive. However, this new system has yet to be tested under a heavy work load, such as that generated in the days just prior to the completion of a major report like the Annual Safety Report or an In-Progress Report.

Given the types of changes affecting ASMIS during the last eight years, it is a credit to its developers that ASMIS is still providing service to the Army accident prevention community. Use of the system, particularly ad hoc requests, have been increasing and are expected to continue to increase. The goal of the redesign of the ASMIS system must be to maximize the productivity of the developers and users of the system by utilizing the available computing facility more effectively.

1.1. BACKGROUND

The US Army Safety Center (USASC) is responsible for the accident prevention activities for the Army. A five-step process is used to improve the accident prevention process. The steps are:

- Information collection. Gather data capable of representing the hazard potential of the entire system.
- Analysis. Define target problems. Discern the systemic causal factors underlying priority problem areas.
- Countermeasures. Use cost beneficial technology to eliminate or reduce risk by improving the system.
- Implementation. Use sound staff concepts to obtain command and staff support for systemic remedies.
- Evaluation. Evaluate the actual performance of actions compared to their projected benefits.

The proper functioning of this process requires data concerning Army accidents. Thus, the USASC is responsible for the collection, computerization and access to data concerning Army accidents.

This accident data is available under the Army Safety Management Information System (ASMIS) which is an umbrella for many databases. Aviation accidents are reported on the DA2397 form and PRAMS and reside in the aviation database. Data is available from January 1972 to the present, with a major form change in October 1983. All other accidents are reported on the DA285 and DA285-1 forms and reside in the ground database. Data is available from July 1974 to the present, with a form revision in October 1980.

All civilian accident claims against the Army are recorded in the Federal Employees Compensation Act (FECA) database which is obtained from the Department of Labor. Data is available from October, 1984 to the present. In addition to the safety related databases, ASMIS houses the Army's Drug and Alcohol database. Data is available from 1981 to the present. This database is transient and will be reassigned to another organization in one to two years. In the future, more safety-related data may be incorporated under the ASMIS umbrella. Possibilities are the Navy Fire Data (fire accident records for all services) and the Night Vision Device Data.

In general, the databases under ASMIS are used independently. The only types of data which provide links from database to database are organization affiliation (e.g., responsible UIC, station, installation) and individual identification (e.g., social security number, name) for each person involved in the accident. The organizational affiliation can be used to access each database and gather information about all accidents for an organization. An example is the number of accidents for an installation for the current year. The identification of an individual can be used to eliminate or link the duplicate recordings in two databases. When an injured civilian files a compensation claim, information is available in the FECA database as well as in either the ground or aviation database. Currently, this link is used to eliminate the duplicate reporting of the accident.

Primary access to the ASMIS databases is through the ASMIS Retrieval and Processing System (ARPS) which provides a query facility accessible to both internal and external users. Unfortunately, the use of ARPS is hampered because:

- The system is difficult to use (not "user friendly"). To become proficient at exploiting the capabilities of ARPS requires time.

- Each database contains a large number of fields and the organization of this data is complex. The ground database contains approximately 200 fields; aviation contains about 2000 fields. To develop a working knowledge of this data requires a substantial investment of time.
- Response to a query can be relatively slow because of heavy use. This was particularly applicable to the IBM 4341 which has since been replaced with an IBM 4381. Slow response will reappear as a problem when the capacity of the 4381 is taxed by an increase in users and/or workload.

Over time, the focus of the internal USASC users has grown from simple reporting of accident statistics (i.e., counts) to analysis of the accident cause(s), and to evaluation of the effectiveness of previous countermeasures. Use of the five step accident prevention process increases ARPS usage and requires in-depth information from the accident report. Because of the delay associated with modifications to the forms (the last modification was in 1983), the narrative portion of each accident report is being used to glean information about recent changes relevant to the cause of the accident or the countermeasures used. This increased use of the narrative places an increased load on the ARPS system and on the internal database users.

The availability of this large historical database has become known outside the Safety Center. Approximately 400 people have been trained in the use of ARPS and the contents of one or more of the databases. The availability of the database, the existence of users trained in its use, and the Army-wide emphasis on safety has increased the number of ad hoc requests from outside the USASC.

ASMIS as it existed on the IBM 4341 (prior to April 1988) had difficulty meeting these demands. From our interviews with the USASC staff in early April 1988, there were approximately 20 TSO users running the ARPS program, and 10 CICS users doing data entry. This load resulted in very slow response. The IBM 4341 was reported to be 98 to 99 percent CPU saturated.

On April 15, the IBM 4341 with 8Mb of memory was replaced by an IBM 4381 with 24Mb of memory. The initial upgrade was strictly hardware and improved system response markedly. The second part of this upgrade (done in the next two to three weeks) involved minor upgrades to some of the IBM software and was transparent to the users. The migration from the IBM 4341 to the 4381 required no change to any USASC software, including ARPS, the COBOL programs used to generate routine reports.

The Army's continual emphasis on safety, the increased use ARPS for safety analyses by both internal and external users, and the increasingly complex information required suggests that to meet future demands, the system must be more available and easier to use for safety analysts outside USASC, and must be more efficient for use by USASC.

1.2. PURPOSE OF THIS STUDY

This study is being undertaken to determine the requirements for an enhanced ASMIS system that can provide effective support to the Army world-wide safety community.

Since the IBM 4341 was replaced just as this study started, this project will study the ASMIS system as it exists on the IBM 4381 and determine what modifications to the ASMIS system can be accommodated on this new hardware.

The focus of this study is on the features and goals of the system and not on the mechanisms used to provide and achieve these. Thus the structure of the computer information system is being collected by interview with the users, programmers, and system manager and not by detailed examination of the code. The current data storage mechanism is being evaluated by review of the file structure and by interviewing the programmers, but only to the level of detail necessary to understand the structure sufficiently to create a model which can be used in the evaluation of various solutions.

1.3. SCOPE OF THIS DOCUMENT

This document is the final report for the requirements analysis of the Army Safety Center's ASMIS system. It includes:

- A description of the current ASMIS system
- The identification of problem areas in the current implementation

- The functional requirements for an enhanced system
- Five proposed modifications to ASMIS system that resolve problems
- A review and ranking of the proposed modifications
- An analysis of the potential problem areas in the chosen modification
- A recommendation for the implementation of an enhanced ASMIS system.

2.0 SUMMARY AND RECOMMENDATIONS

This chapter presents the recommendations and a brief summary of the work leading to the recommendations. Detailed explanations of this work are available in Chapters 3 through 6 of this document.

2.1. SUMMARY

The existing ASMIS system was reviewed and documented (see Chapter 3). From our understanding of this system and interviews with the USASC programmers, and both internal and external users, functional requirements for the enhanced ASMIS system were developed. To aid in maintaining continuity between the existing and the enhanced systems, the features and problems of the existing system were associated with each functional requirement (see Chapter 4).

Five possible implementations for the enhanced system were developed. These implementations were compared using the functional requirements and administrative issues, such as cost of purchased software, level of training required and cost of implementation as metrics (see Chapter 5). The implementation based on a database management system (DBMS) was rated highest.

A number of other issues were investigated. The utilization of the IBM 4381, the style and number of ARPS queries, and the current size and growth rate of the ASMIS databases were investigated to further describe the conditions under which the enhanced ASMIS system must function. To evaluate the appropriateness of the DBMS-based solution, the effect of using a DBMS instead of the current ARPS program to respond to queries was modelled. To identify the best DBMS, the available systems were identified and investigated.

The first subsection describes the highlights of these investigations. Full descriptions are available in Chapter 6. The final subsection consolidates the results of the investigations and describes the scenario for the enhanced ASMIS system.

2.1.1. Results of Investigations

Evaluation of the performance of the IBM 4381 indicated that the system is approximately 34 percent utilized. Thus the 4381 could comfortably support double this load, but tripling the load would saturate the system (see Section 6.1). This implies that the Centralized Multiple User Query Processor (Section 5.3) and the DBMS with Multiple User Query Processor (Section 5.5) are inappropriate solutions.

The evaluation of the ARPS queries was done using a log of approximately two weeks of queries. The ground and aviation data are responsible for 91 percent of these queries. In these two databases, there are two different query styles. The first style, non-matrix queries, produces one line of output for each database case selected (exclusive of narrative). This style is characterized by a large number of short requests (approximately 85 percent require less than 150 database cases) and a few large requests (less than 5 percent require more than 1000 cases). The second style, matrix queries, produces a table summarizing the relationship between two fields (multiple tables can be generated to summarize three fields). A much larger number of database cases are retrieved: an average of 1521 database cases for the ground database and 777 for aviation (see Sections 6.6.1.1 and 6.6.2.1).

Comparison of the functionality desired for the enhanced ASMIS system with the standard data models resulted in choosing an information systems environment. The appropriate database model for this environment is a relational database (see Section 6.2).

Models for a relational DBMS and for the current ARPS program were developed to evaluate the effect of using a relational DBMS instead of the current ARPS program. Comparison of the predicted ARPS response and the predicted database response to the total collection of queries indicated a difference. Non-matrix queries can be handled by the database model as efficiently as ARPS does. The evaluation of the matrix queries indicated a difference; the database would respond more slowly. The database would respond 2.2 times slower for ground data (3.7 times slower for the matrix queries) and 1.5 times slower for aviation (2 times slower for the matrix queries).

A more complicated model of the relational DBMS, involving structuring the storage of the data, was proposed. The predicted database response to the total collection of queries was recalculated using this new model. Comparison of the predicted ARPS response and the predicted database response indicated that the collection of all queries could be handled by the database model as efficiently as by ARPS (ground would be 1.04 and aviation would be 0.98 of the ARPS time).

Elimination of repetitious retrieval of database cases, particularly for matrix queries, could improve performance. A review of the session style indicated that some sessions were using the ARPS program to repeatedly select the same set of cases (or a subset of those cases) and generate a single matrix. A more efficient use of computer resources would be to generate multiple matrices from the same subset of data and avoid reselecting the database cases. Detailed examination of only six sessions from the ground database resulted in reducing the number of times the database must be accessed for matrix queries by 43 percent (136 accesses to the database were replaced to 6 accesses and the generation of multiple matrices from those sets of data). The DBMS-based system would now take only 66 percent of the time required by ARPS. A similar analysis reduced the number of times the database would be accessed for aviation matrix queries by 16 percent and the DBMS-based system would then take 87 percent of the time required by ARPS.

Because of the high percentage of repetitious work caused by the matrix queries of just a few users, all queries were analyzed to determine how much repetitious work was being done. Overall, 35 percent of the ground database and 26 percent of the aviation database queries indicated reuse of the previously selected set. Considering only the queries where the user actually had a previously selected set revealed that 62 percent of the ground and 48 percent of the aviation queries could be satisfied with no reselection of database cases. Because of the current functioning of ARPS, the desire to reuse the previously selected cases does not imply that the data would be reused exactly. An additional selection criteria to further limit the set could be added by the user. An example of the type of user session we observed will clarify the required capability. The user selects all helicopter accidents for the last five years. To compare the experience of his installation with that of the Army as a whole, he does a series of matrices for the whole Army and a companion series for his installation. Thus, the process of creating multiple matrices or multiple reports from the same set of data would need to be able to do a simple selection (Sections 6.6.1 and 6.6.2 include other examples).

Noisy communication lines and loss of carrier on these lines also causes repeated work. If the noise adds characters to the user's query, he must repeat the query to get the desired results. If carrier is lost, the user must log in again. If he reestablishes the connection within the six minute interval, his job continues; otherwise, the query must be resubmitted (see Section 4.4.5).

Use of a DBMS requires that data under its control be on random-access devices. The maintenance of historical data on a sequential medium (tape) presents an implementation problem. Two access methods, one for the on-line data and one for the off-line data, would be necessary (see Section 5.4.3). Projected data storage requirement, including all current on- and off-line data, is 30 to 45 percent of the current 7.5 gigabytes of disk storage. Adding another 2.5 gigabyte disk drive results in the data occupying 22 to 33 percent of the storage space. In the 10 gigabyte configuration, the data storage requirement for the current databases would increase at approximately 2.3 to 3.4 percent per year (see Section 6.7). This would provide sufficient storage for a DBMS-based ASMIS system for three to four years. Thus purchase of disk drives, for approximately $60,000 each, is more cost effective than design, implementation and maintenance of an additional access method.

In the future, optical disk technology may provide an alternative to the current magnetic media. An article, "Optical Storage Comes of Age," (Levine, 1988) provides insight into the current state of optical disk technology. For ASMIS to effectively use optical medium to store older data two things must happen. First, an optical disk must be available to add to the IBM 4381. Second, the DBMS software, DATACOM/DB, must support the use of the optical disk.

Five relational database products were identified based on their ability to run on an IBM 4381 under the MVS/SP operating system. Of this initial list, three DBMS were eliminated from consideration on technical grounds or their uncertain future. The remaining two DBMS packages, DATACOM/DB and SUPRA, were evaluated and found to satisfy the requirements of ASMIS. However, the cost of the systems to the Army Safety Center is significantly different. The Army has purchased a site license for DATACOM/DB and related products which will allow the Army Safety Center to acquire DATACOM/DB at no cost. SUPRA, on the other hand, would have to be purchased for approximately $269,000.00. It is recommended that the Army Safety Center obtain DATACOM/DB and its related products as outlined in Section 6.7.

Currently the Army Safety Center's IBM 4381 is running the MVS/SP operating system. However, there are two newer versions of MVS available, MVS/XA and MVS/ESA. Since 1983, IBM has made no significant extensions to MVS/SP, but instead has concentrated its efforts on MVS/XA and its successor, MVS/ESA. Unfortunately, MVS/ESA only runs on the E series of mainframes. The Army Safety Center's IBM 4381 MG13 would have to undergo a major hardware upgrade to the model group 91E or 92E in order to run MVS/ESA. However, an upgrade to MVS/XA under the current hardware configuration is possible. MVS/XA would allow ASMIS to take advantage of current MVS enhancements, and eliminate the possibility of encountering the address space limitation under MVS/SP. This upgrade also changes the cost of the operating system from an ongoing lease to a 36-month payment plan.

2.1.2. A Scenario for the Enhanced ASMIS System

With thorough analysis and careful design, an enhanced ASMIS system based on a database management system could do what ARPS currently does plus more. The creation of this system would require a significant amount of training for the programming staff. Effective use of the enhanced system would require training for the users, particularly the USASC staff members in the RAID and SMD groups. Long-term functionality would require knowledgeable tuning of the database and careful restructuring to reflect the changes arising from changes in the Army safety community.

A critical feature of the careful implementation would be the use of the knowledge that a substantial portion of queries are based on the data selected by the previous query. This knowledge could be used to eliminate repetitive selection of database cases and implies that access to the Army safety data would be a two step process. The first step selects the data from the database. The second step produces reports or summarizations from the selected subset of data. To provide this two step structure, the environment should provide the following options:

- The user should be able to transfer a set of data from the mainframe to a personal computer. The transfer of this data set involves not only moving the data from machine to machine, but also the movement of the field names and code translations plus the incorporation of this information into an analysis or reporting package. This whole operation must be very easy for the user.
- The generation of matrices and other statistical summarizations should be removed from the database management system and handled by a statistical analysis package. This kind of a package should be available on both the mainframe and on personal computers.
- Multiple reports should be generated from the same set of database cases. This function may be a part of the database management system or it may handled by another package. This functionality should be available on both the mainframe and on personal computers.
- On the IBM 4381, the user should be able to save the data selected on disk and return later to generate additional matrices and reports.
- On a personal computer, the user should be able to save the data transferred from the IBM 4381 on disk and return later to generate additional matrices and reports.

2.2. RECOMMENDATIONS

The enhanced ASMIS system should be a DBMS-based implementation with three important features; elimination of repetitious work, flexibility and ease of use. These features will be obtained by careful planning and implementation using a variety of software tools on both the IBM 4381 and personal computers. Figure 2.0 is a high level diagram of this structure which includes the following:

- A central database using database management system (DBMS) to provide data organization and selection.
- A statistical analysis package to provide the summarizations.
- A reporting package to provide the columnar listing and caseprints.
- The IBM 4381 will be the center of the system. It will provide storage for the database. All data in the database would be stored on-line. It will also provide storage for data subsets selected by users.
- For USASC staff members and other frequent users, personal computers could be analysis stations, thus offloading much of the analysis and reporting work and data subset storage from the IBM 4381.
- For infrequent users and those without personal computers, the IBM 4381 will provide the same analysis and reporting functions and disk storage for data subsets.

To eliminate repetitious work by removing the reselection of database cases requires a new paradigm for the system. The paradigm will be the selection of a subset of data by the DBMS followed by the generation of multiple summarizations, columnar listings and/or caseprints by the statistical analysis and reporting software.

Flexibility will be provided by using a database management system to store the data and to select a subset of database cases specified by a query. This will provide data independence so that changes to the structure and content of the data can be made with minimal changes to existing applications.

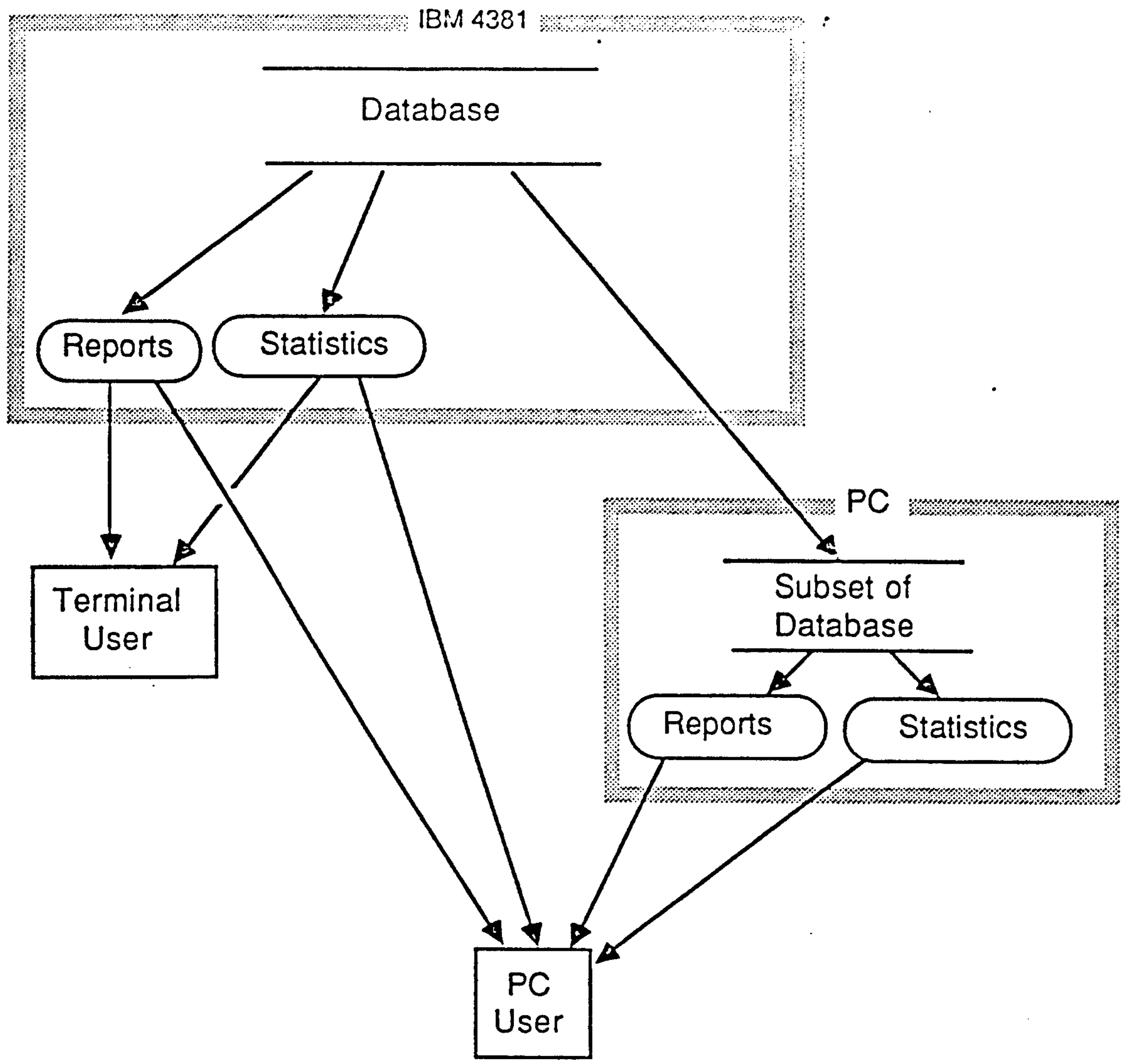


FIGURE 2.0 Conceptual View of Proposed System

PC users can download data to the PC and generate reports and statistics. PC users can also use the PC as a terminal. The terminal user does all work on the IBM 4381.

An important part of ease of use can only be achieved through user education. The USASC must realign its DOIM staff to support the dual computer environment by providing experts for both the IBM 4381 and personal computer software. It must also be realigned to support the multiple package environment by providing experts for the DBMS, the statistical analysis package, and the reporting package. Most important of all, these experts must interact with the users, both USASC staff and external users to provide initial and on-going training, assistance and problem resolution. A major goal for this interaction is the elimination of repetitious work, by helping the user structure his requests so that multiple reports and summarizations can be generated from the single data subset.

2.2.1. Computer Software

The software recommendations include:

- The database management system resident on the IBM 4381 should be DATACOM/DB. Through an agreement between the US Army and Applied Data Research, DATACOM/DB can be acquired at no cost.
- The statistical analysis package for both the IBM 4381 and the personal computers could be SAS. SAS Institute, Inc. is developing direct access to data stored in a DATACOM/DB database from within SAS. January 1989 is the target date for a beta test version with a production version anticipated during 1989. A definite choice will depend on a more detailed study of the needs of the users, particularly the frequent users, and on the applicability and progress of this interface.
- The reporting package could be either a package associated with DATACOM/DB or SAS. The choice will depend on a more detailed study of the needs of the users.
- The software for the personal computers will also depend on the needs of the users and thus requires more study. A variety of personal computer software can be used, including: 1) a statistical analysis package, 2) a reporting package, and 3) a terminal emulation package.

- MVS/XA should be obtained. MVS/XA is not a requirement for the initial development of the DBMS-based system. Because IBM is no longer making improvements and extensions to MVS/SP, MVS/XA would provide some improvements currently and eventually will be necessary to gain access to new features.

2.2.2. Disk Storage

All data in the ASMIS database should be maintained on-line. Users will require space for storage of data subsets. To provide adequate space an addition 2.5 gigabyte disk drive should be purchased. If no substantial reduction in the 33 percent of the current disk space unavailable for data storage can be accomplished, adequate space can be obtained by purchasing two 2.5 gigabyte disk drives. The approximate cost of a disk drive is $60,000. The purchase of the second disk drive may require the purchase of a disk controller also.

2.2.3. Organizational Structure

Within the DOIM structure there should be three identified positions; a database administrator (DBA), a statistical analysis expert and a personal computer expert. A backup person for the DBA should also be identified.

The DBA is the central person in the use of the database management system. The DBA is responsible for:

- Maintenance and enhancement of the database. The DBA analyzes the overall implications of proposed changes and coordinates the actual changes.
- Maintenance of a document which is a high level description of the database. This document includes descriptions of the database structure, all data fields, all code translations. It is updated to reflect changes and documents the date and reason for such changes. Two types of computer software are available to assist in the maintenance of this document. A CASE (computer aided software engineering) tool can be used to maintain the high level description of the database. CASE tools are available for personal computers. Costs for an IBM PC-based package range from $2,000 to $10,000. A change control product can be used to maintain a

record of the changes made to the database structure,. fields, code translations, etc. PANVALET, set to save all version of a file, could be used.

- Database Backup. The DBA is responsible for having the database backed up and for the maintenance of adequate archival versions of that backup.
- Data security and the granting of user privilege to access the data within the database.
- Monitoring of database performance and tuning of the DBMS.

The database administrator will be a key person in the success of the DBMS-based ASMIS system. The existence of a backup person who actively participates in the database administration activities and a written description of the database itself will reduce the dependence of the USASC on this person's expertise.

The statistical analysis expert knows the statistical tools picked for the new ASMIS implementation and assists users in their use. This expert provides initial training and continual support, guidance and problem resolution. This involves continual interaction with the users, both inside and outside the Safety Center. Moving the statistical analysis work out of the DBMS and into a package which more efficiently generates the necessary statistics and summarizations is necessary for the success of a DBMS-based ASMIS system.

The personal computer expert knows the PC-based products chosen for the new ASMIS implementation and assists users in their use and acquisition. This expert provides initial training and continual support, guidance and problem resolution for the users. This involves continual interaction with the users, both inside and outside the Safety Center. Moving much of the analysis and reporting work from the mainframe to personal computers will enhance the functionality of the DBMS-based ASMIS system.

2.2.4. Training Requirements

Training of both the data processing staff and the users of the database is required for success of the DBMS-based implementation of ASMIS. Lack of user training will result in no change to the user-ARPS interaction style. This style is repetitious extraction of database cases and will result in the saturation of the IBM 4381. Lack of training for the programming staff will result in inefficient organization of data and ineffective use of the database management system. In turn, this inefficiency will lead to saturation of the IBM 4381. Each member of the data processing staff needs training to gain knowledge of the structure and functionality of the database management system.

The database administrator and the DBA backup need training in relational database technology. This requires training in the structure of relational databases, the specific structures used by the chosen DBMS, and on the creation, maintenance and tuning of databases under the chosen DBMS. This can be gained from a vendor-independent course on the relational database model plus intensive training from Applied Data Research, Inc., the vendor of DATACOM/DB. The DBA and his backup should be active in the user's group for DATACOM/DB as a source of continual learning and technical interchange with other shops.
Section 6.5 provides more specific information on the courses available from DATACOM.

The statistical analysis expert needs training in the statistical packages available on both the IBM 4381 and on the personal computers. Like the DBA, he needs to be active in the user's group for these packages as a source of continual learning and technical interchange.

The personal computer expert needs training in the PC packages chosen as part of the new implementation of ASMIS. He needs to maintain an awareness of new products. Attendance at a yearly trade show for PC-based software would provide an awareness of new features in the chosen products and an awareness of new packages.

All members of the ADP programming staff who will be using the DBMS should receive training appropriate to the modules of the DBMS system they use. Section 6.5 provides more specific information.

For the ASMIS users, training is required in the methods of selecting, manipulating and analyzing data. In particular, users will need to understand the "select the data once," followed by "generate multiple reports or summarizations" organization of the processing. Users will not need to understand the theory of the DBMS. They will need to understand the organization and content of the data and the analysis tools available to them. A three-pronged strategy will provide the necessary training. First, an initial training on the use of the chosen packages and their application to the Army safety data is necessary. Second, continual self-learning should be facilitated by improved user documentation, on-line help and development of a facility to exchange problems and their solutions. Third, the USASC users should form a user's group (USASC users are approximately 70 percent of the ARPS user base, see Section 6.6.4). Regular meetings will provide a verbal interchange of techniques and will provide a forum for presenting problems to the programming staff. The personal computer expert, the statistical expert and the DBA should attend these meetings, both to provide solutions to current problems and to learn of anticipated areas of emphasis and associated areas of user concern.

2.2.5. Structure of the New ASMIS Implementation

Like the current implementation of the ASMIS system, the enhanced system will have four components: the ad hoc query facility (ARPS), the routine reporting facility, a data entry facility and a facility to precompute information during times of low system utilization. Each of these facilities will be contain one or more application programs.

A high level document describing these applications and their interrelationship should be created and maintained. The use of a CASE tool would facilitate this activity.

Change control should be applied to these applications. This provides a history of the changes to all applications and makes identification of programs which will be affected by a database change much easier. Using this method, the programs which need change (or at least those to considered) could be identified by a simple computer run.

2.2.5.1. The Ad Hoc Query Facility (ARPS)

The specification and development of a DBMS-based ARPS replacement requires further analysis. The following items should be included in the design:

- The ARPS replacement should have: 1) a DBMS interface to select the database cases requested, 2) a statistical analysis package to produce summaries of the data (the current matrix output), 3) a reporting package to produce the columnar listings (the current non-matrix output) and caseprints, and 4) the ability to store sets of data on disk so the user can return later and generate more reports or summaries. Ideally the generation of summaries, reports and caseprints should be done by the same program, but this may not be possible. SAS with its direct interface to data in DATACOM/DB databases would allow these facilities in one program. It deserves close scrutiny as the user interface.
- The ARPS replacement should be structured so it can be run partially on personal computers or all on the IBM 4381. The selection of data from the database would be done on the IBM 4381. The generation of reports, caseprints or summarizations could be done on either the IBM 4381 or a personal computer. The storage of selected sets of data should be possible on either the IBM 4381 or a personal computer. SAS functions on both the IBM 4381 and personal computers and thus could provide one user interface which works on both the mainframe and personal computers.
- Feedback should be available to help the users in developing queries which are easily and quickly fulfilled by the DBMS. In essence this requires an overview of the contents of the database so the user can make educated choices in the specification of his selection criteria.
- On-line help should be included in all parts of the enhanced ARPS. This help should be organized so that it can be successively disclosed to the user. The on-line help should include: 1) what queries have precomputed results and how to display that information, 2) instructions for selection

of data, 3) a link into the centralized query library, 4) an overview of the organization of the database, 5) for each field, a definition and a list of all codes used, including the date the code was added or removed from use, 6) a list of the changes to fields and codes to facilitate users changing their queries and analyses to reflect the database changes, 7) instructions for creating columnar listings, 8) instructions for creating caseprints, and 9) instructions for creating summarizations.

- Continued monitoring of the query patterns should be part of the ARPS replacement. The queries submitted should be monitored so that frequently used queries can be identified and added to the set of queries that is run overnight.

- The ARPS replacement could be a screen-oriented program (in contrast to its current line mode operation). The use of SIM 3278/VTAM on the IBM 4381 allows most CRT terminals to function as IBM 3278 devices. To make screen-oriented operation using SIM 3278/VTAM transparent across a wide variety of CRT terminals, function keys should not be used because the mapping of IBM 3278 function keys onto the user's keyboard varies from terminal to terminal.

  Making ARPS be a screen-oriented program would mean that line mode devices such as a TTY or a Tektronix Silent 700, could not be used. The number of non-CRT devices used to access ASMIS is unknown, thus no definitive recommendation can be made.

- The ARPS replacement should automate the translation of all codes. The user should be able to easily specify the use of a field or its translation in his reports and summarizations. The translation of the codes available in the printed documentation should be used only as a reference; users should not be required to use it as the means of translating codes.

Currently ARPS provides ad hoc query access to the ASMIS databases for both the USASC ADP staff and all other users. Thus, it is used by both experienced and naive users. In the enhanced system, two methods of access to the ASMIS databases should be available. One access method should be available for use by naive users and require little training. A second, more powerful method should be available for users who are willing to spend time in training. For these experienced users, a structured query language such as ADR/DATAQUERY or ADR/DATAREPORTER should be available. This provides a powerful, but simple to use (as compared to writing COBOL) access method for the user knowledgeable in the products use and in the structure and contents of the ASMIS database. For naive users, who does not have the knowledge of the structure and contents of that database, a guided step-by-step interface tailored for ASMIS and the database contents should be available.

Generation of this tailored interface for naive users does not mean that existing packages, such as SAS should not be used. One possible way to use existing programs would be to have the ARPS replacement program elicit the user's request, obtain the data from the DBMS and create command files to drive the other tools chosen for ASMIS. Other implementations also exist.

By providing a well documented structured query language processor such as ADR/DATAQUERY or ADR/DATAREPORTER, USASC may find that some internal users, such as those in RAID and SMD, will become knowledgeable users. This would provide two benefits. First, it would reduce the level of sophistication necessary in the ARPS replacement program. Second, it would reduce the routine work load of the ADP staff and allow them to concentrate on improvements and modifications to the ASMIS system.

2.2.5.2. The Other Application Programs

The programs for routine reporting and other necessary applications should be developed within the tools accompanying DATACOM/DB or by using the ARPS-replacement program. Many of the routine reports can be generated using these two approaches, but some complex reports may still require the use of COBOL. The program generator available with DATACOM/DB may be helpful in the development and maintenance of these programs.

2.2.5.3. Data Entry

An on-line data entry system needs to be implemented in which changes are reflected in the DBMS as soon as they are made. The enhanced ASMIS information system eliminates the need for delayed batch updating by allowing a data set to be extracted and multiple operations to be performed on the data set. Real-time updating will give analysts immediate access to data as soon as it is entered and provide greater accessibility to data. It will no longer be necessary for ARPS to be down for nightly updates. If verification is needed before the record should be used for analysis, flags may be added to the database structure to indicate validated records (normal users would have access to only validated data). The DBMS logging facilities should be activated to ensure that the daily data entry updates will not be lost in case of database corruption. Other benefits of online data entry include:

- Less complicated DBMS data entry implementation. By combining the data entry process and the batch data file updating process into one, there is less code to keep track of and batch scheduling is not needed.
- Elimination of blind updates. This will provide better data entry verification by allowing the data entry clerk to view the existing record before updating the record.

2.2.5.4. Overnight Processing

Overnight processing should continue to be used for the generation of routine reports. Generation of the frequently used statistics (the statistics data option of ARPS) should continue to be done each night. The usage of this statistics data option should be monitored to determine which of the calculated data are being used. Ideally, the pattern of database queries should be monitored to determine when the frequently used statistics should be supplemented to reflect the change in the needs of the user community. In practice, this may be very difficult to automate and the determination of what to add may be done by hand.

2.2.6. Other Recommendations

2.2.6.1. System Security

The current need-to-know and periodic revalidation of that need should be maintained. USASC should purchase and install an access control system such as RACF, ACF2 or Top Secret. This product should be used to protect data and programs from intentional or unintentional destruction or corruption by establishing appropriate protections on all existing files. The system should be modified to establish appropriate protections on newly created files.

Each user should exist in a captive environment. He should have access to the necessary programs (the ARPS replacement and the statistical analysis and reporting packages) and his own data files. He specifically should not have access to the files of another user.

USASC should install an on-line password facility. A user should be able to change his password immediately without contacting USASC.

The article entitled "Stalking the Wiley Hacker" (Stoll, 1988) describes the attempts to track and identify a single hacker. It also provides valuable insight into the security aspects of a computer system attached to an network.

2.2.6.2. System and Data Backup

The regular backup, including incrementals between full backups should be continued. Adequate versions of the backups should be maintained. The use of off-site storage for both system and data backups should be reinstituted.

2.2.6.3. User Documentation

The documentation for the database, as presented from the perspective of the user, should include:

- An organizational overview of the database, describing the parts of the database, its contents and the interrelationships. It should also describe the multiple versions of one type of data, for instance the current aviation data and the pre-1972 aviation data.
- Detailed information about each field: a definition, all codes used, the date of any coding changes. To help users understand form changes, the information for each field should contain a pointer to predecessor and successor fields as appropriate.
- Specific instructions for the use of the various parts of the new ASMIS system. In particular it should include step-by-step instructions for

transfer of data between ARPS and the statistical programs, both on the mainframe and on personal computers.

On-line documentation should be integrated into the various parts of the ASMIS system.

2.2.6.4. Data Communications

USASC should continue to provide access to the ASMIS system for as many hours a day as possible. This facilities world-wide use of the database.

USASC should provide 2400 baud dial-in communications as well as the existing 1200 and 300 baud.

Some problems exist with the dial-in communications. Often the communication line is noisy. Occasionally the carrier is dropped. Two alternatives are available to provide a checked transmission protocol. Either of these two options will handle occasional noise on the line and neither will be effective on a very noisy line. MNP modems can be used independent of the device used as a terminal (either an actual terminal or a personal computer). MNP is Microcom Networking Protocol and uses hardware to detect incorrect transmission and retransmit incorrect messages. To allow use of MNP modems, both the user and the USASC must have MNP modems. The cost of a 2400-baud MNP modem is in the $500 to $700 range. For users with personal computers, the SIMPC package can be used. This requires that SIM 3278/VTAM be run on the IBM 4381. The SIMPC - SIM 3278/VTAM pair use software to detect incorrect transmission. The cost of SIMPC is approximately $250 per machine; SIM 3278/VTAM is currently available on the IBM 4381.

3.0 THE EXISTING SYSTEM

For purposes of this description, the current system has been broken into eight components which will be discussed separately. The components are:

1) The data flow between the ASMIS system and its users, both internal and external to the Army Safety Center (USASC)
2) The databases available under the ASMIS system
3) The routine processing done by USASC
4) The access methods used to respond to user requests
5) The data analysis facilities available to users
6) The available user documentation
7) The security mechanisms in use
8) The computer hardware and software.

To convey the significant details of the current system, both text and diagrams will be used. For the diagrams, icons have been used to represent the following types of structures. Figure 3.0 shows a sample of each icon.

- Organizational entities are groups of people. Examples are USASC, Office of Information Management (DOIM), and Office of Research, Analysis and Investigation (RAID).
- Computerized entities are structures which exist on a computer. Examples are the ASMIS system, the nightly process of updating the ground and aviation databases and the processing necessary to produce the safety goal numbers.
- Stores of data. Typically this is a file on a computer.
- Lines with arrows connect organizational entities, computerized entities and data stores. The arrow head indicates the direction of information transfer and the labelling explains what is being transmitted. Because of space limitations, lines with an arrow at each end indicate

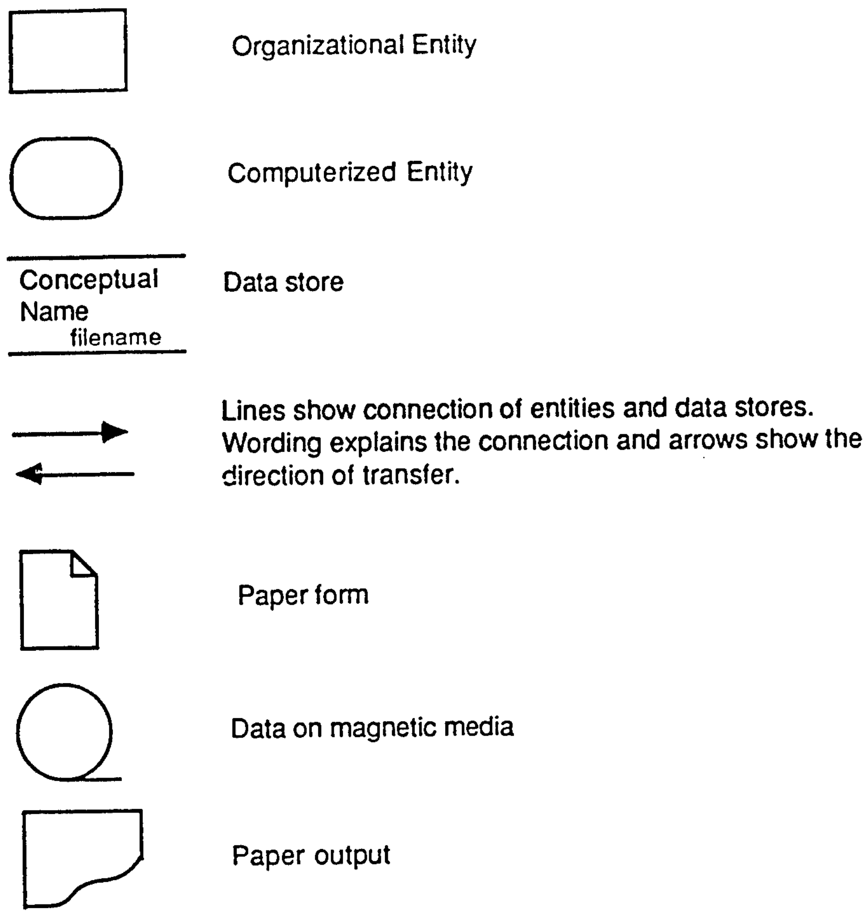


FIGURE 3.0 Legend for other figures

the transmission of two types of data. For example, a single line indicates the transmission of a query to the Safety Center and the return of a report to the external user.

- Paper forms. Examples are the DA285 and DA2397.
- Data on magnetic media. An example is the flying hours tape.
- Paper output. An example is the update error list generated by the nightly update of the ground and aviation databases.

3.1. DATA FLOW

Data accessible under ASMIS comes from a number of external sources and is used by a variety of users both inside and outside the USASC. The ASMIS system will be described from the perspective of its interface to the users and data suppliers external to the Safety Center and its use within the Safety Center.

3.1.1. Interfaces External to the Safety Center

Figure 3.1 depicts the relationship of the USASC to the data suppliers and database users outside the Safety Center. All of these interfaces require human interaction, except possibly the interface to the external user.

There are three styles of data input: electronic transfer, paper copy and computer tapes. Except for the Drug and Alcohol data, the entry of data from paper copy is handled by the Accident Records Management Division. The computer tapes are handled by the Data Processing Division. For the ground database, the DA285 form and the exposure report (DA2398) are received on paper or via electronic transfer or diskette. For the aviation database, the DA2397 is received on paper and the flying hours are on tape. For the FECA database, both the monthly and quarterly data are received on tape. For the Drug and Alcohol database, all information (DA4665, DA4666, DA3711-R and DD2398) are received on paper and entered by the U.S.A. Drug and Alcohol Operations Activity (USADAOA) in Washington, D.C.

There are three basic data request styles available to external users. A user may use a terminal and directly access the ASMIS system; he may contact

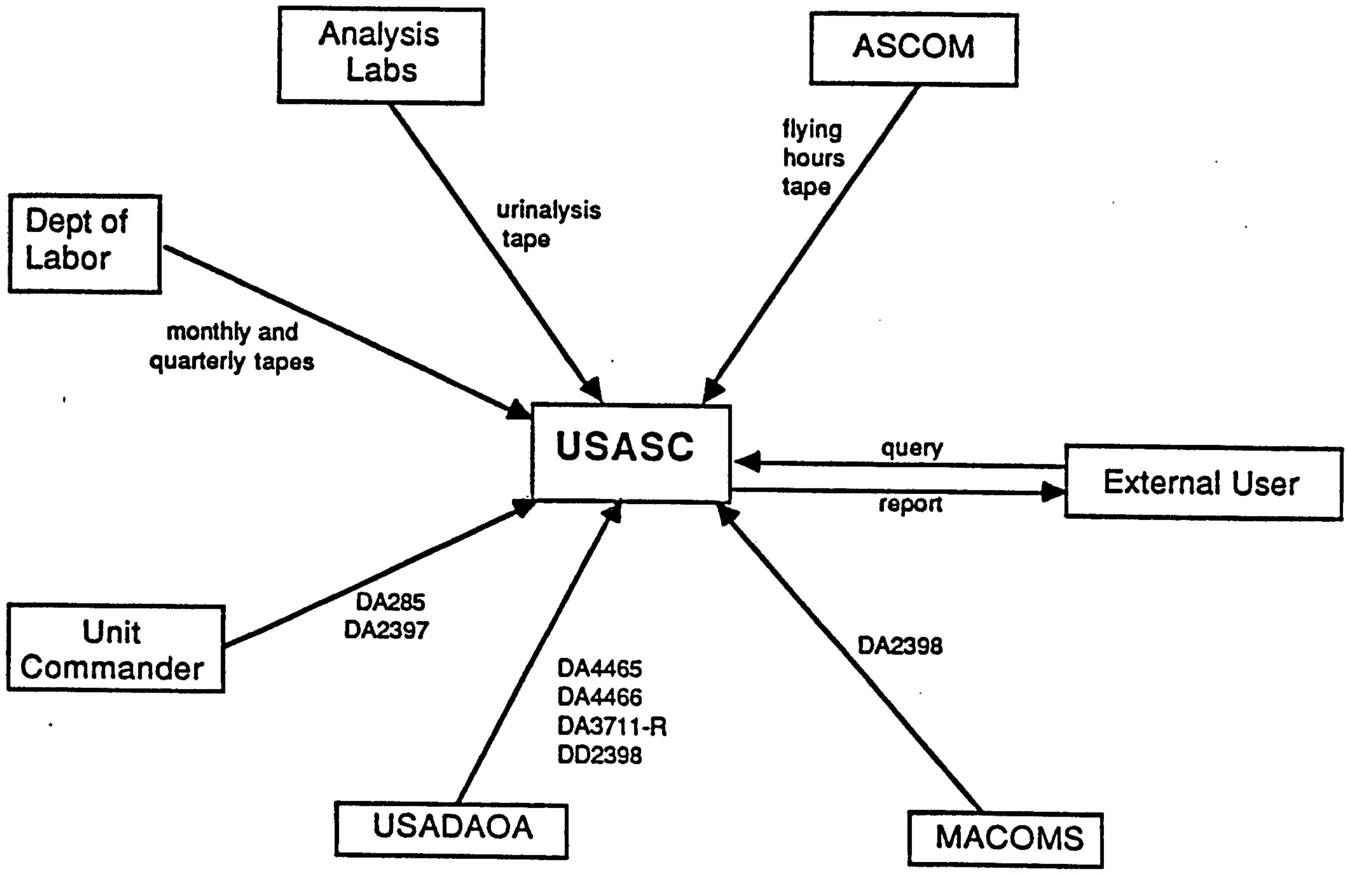


FIGURE 3.1 USASC External Interfaces

DOIM (phone calls, electronic mail, formal and informal written correspondenceare used) to have his one-time request fulfilled or he may ask that an answer to his request be produced routinely (e.g., monthly, quarterly).

The external user may, if he is trained and has the necessary access equipment, log onto the computer system and using the ARPS program, access the databases in the ASMIS system directly.

Users who have either a one time need or lack the necessary computer equipment or training, contact the Safety Center and have their ad hoc requests fulfilled by someone in DOIM. For this type of query, either ARPS, SAS or a COBOL program is used. For simple requests, ARPS is used; for more complex requests SAS or a COBOL program is written.

For users with recurring requests, the method for answering the request is saved and the running of the request is incorporated into the routine requests system where it is automatically executed on the desired schedule (e.g., monthly, quarterly). Examples of this type of request are the quarterly Aviation Case Prints sent to aircraft manufacturers, the quarterly National Stock Number Report (aviation) and the quarterly National Guard Accident Experience Report (ground). A few of these requests require user interaction to produce the desired reports. An example is the quarterly Selected Pieces of Materiel Report sent to CECOM, which requires a list of equipment as input to the report process. Not all recurring requests are scheduled; there are a small number of reports which are requested by users on an irregular basis.

Individual data requests may move from one category to another. For instance, a one time query can become a routine request. A user who contacts DOIM to fulfill his request may eventually desire training and use the ASMIS system directly.

3.1.2. Interfaces Internal to the Safety Center

Figure 3.2 depicts the structure of the Safety Center as it relates to ASMIS. The USASC personnel associated with or using ASMIS are identified by their division acronyms and their interface to the computerized data system is shown.

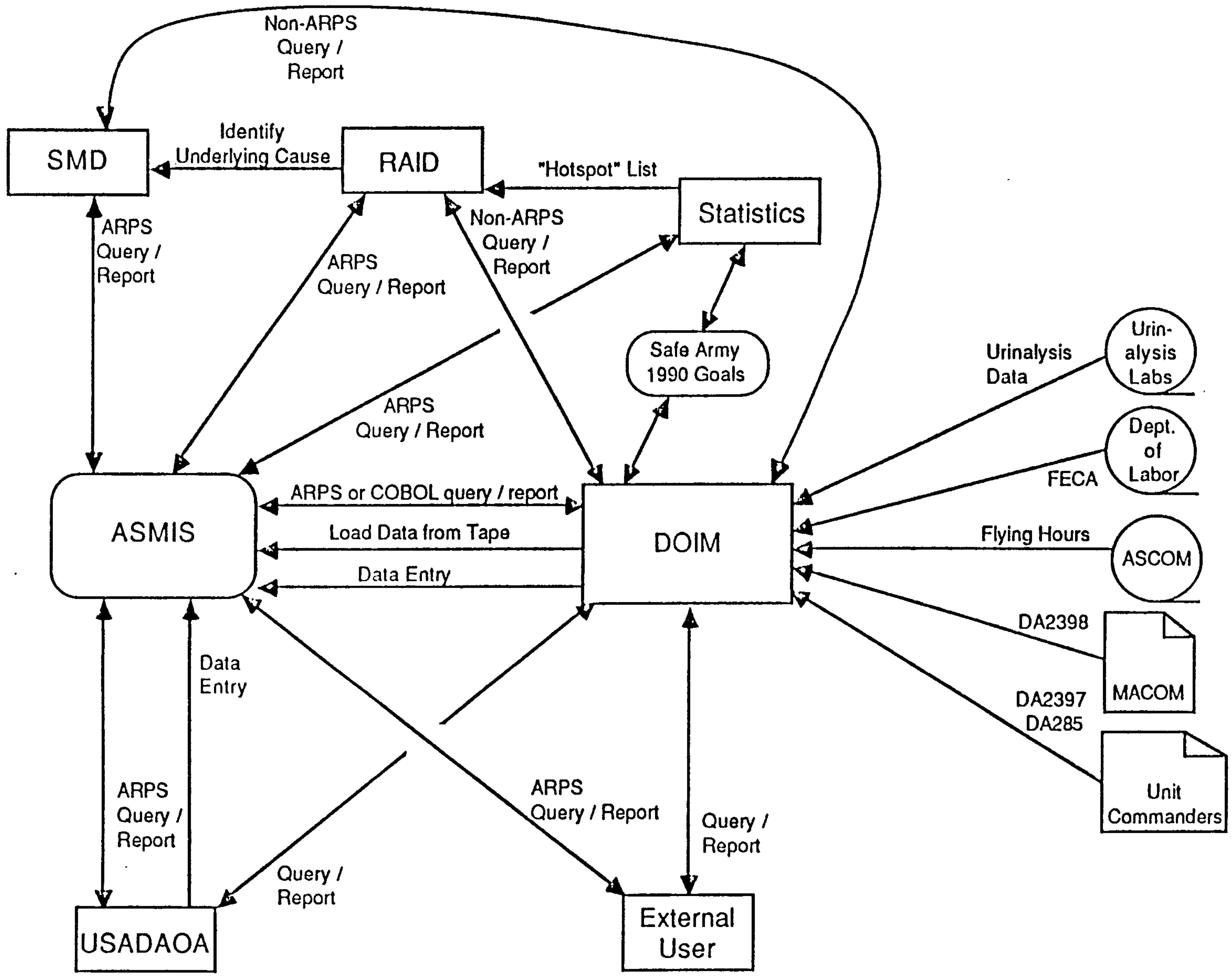


FIGURE 3.2 USASC Internal Interfaces

There are four groups within the USASC who have regular need to access the databases within ASMIS. Data is supplied to other USASC directorates as a function of the access of these four groups.

Addition, correction and deletion to all databases, except Drug and Alcohol, are the responsibility of DOIM. As shown on Figure 3.1, some data is entered from paper copy and some is transferred from magnetic tape. In addition, DOIM is responsible for answering queries coming from the external users. This includes both ad hoc queries and generation of routine reports.

The statistics group is responsible for the generation of the Safe Army 1990 information. This requires requesting data from DOIM to use in the accident estimation process and produces the estimate of current accident rates for the fiscal year and the Safe Army 1990 goals. Monthly, the accident rates and goal numbers are formatted by DOIM and made available under the Statistics option of ARPS. This information is published quarterly.

The statistics group also produces a "hotspot" list of accident types. The purpose of the "hotspot" list is to identify accident types which should be investigated to see if modification to training or equipment can reduce the number of accidents.

RAID is responsible for analyzing accident types to determine the systematic underlying cause(s). The source of the accident types to be investigated is the "hotspot" list plus selected requests from outside the USASC. Because of the cyclic nature of the accident prevention process[a], and the relatively static nature of the input forms (DA285 revision date is August 1980, DA2397 is March 1983), access to the narrative data is often necessary. For most accidents, this information is available through ARPS. For some severe ground accidents, this requires data from the 285-1 form which is not currently available under ARPS. The 285-1 data is accessed through a COBOL program and requires the assistance of a DOIM programmer.

(a) The accident prevention process includes analysis of the accident type, development of countermeasures, implementation of these measures and checking on the use of the countermeasures during the analysis of new accidents of the same type.

Office of Systems Management (SMD) is responsible for determining how to improve the system to reduce the risk of accident (e.g., revise training instructions and correct materiel defects). This also requires detailed information from the accident report and thus the type of data needed and the mechanisms for data requests are the same as used by RAID.

3.2. DATABASES

The Army Safety Management Information System (ASMIS) is an umbrella for a number of databases. Currently each database consists of a number of data files. Some of these files are accessible using the ASMIS Retrieval and Processing System (ARPS) and some are not.

For purposes of this system requirements evaluation, the ASMIS system includes the aviation accident database, the ground accident database and the civilian accident (FECA) database. These three databases will be used to model the structure of the proposed system.

The Drug and Alcohol database is transient and is expected to be removed from the Safety Center in one to two years. It will be included in the analysis, but in a limited fashion. This database will not be used to determine the structure of the proposed system. Rather, it will be used to indicate the increased capacity necessary to allow the inclusion of this or other transient databases.

On the USASC computer system there are a number of other databases which will not be included in this analysis. Examples of these are the TDY data, the publication and film lists, and the various mailing lists. These databases are small and of well-known functionality. They can be left "as is," added to the new system, or moved to a personal computer.

Each section below is a summary description of one database. The following items are included in each description:

- Database Contents. This is a description of the conceptual contents of the database and links the conceptual pieces to actual data files.

off-line records and the growth rate. On-line records are those stored on the disk. Off-line records are stored on magnetic media.

- Auxiliary files. These are files which contain information about the database and are used by the ARPS program. A data dictionary describes the record (e.g., the aviation record) as it appears inside the ARPS program and is used to select records and format reports. The code book is used to translate from the code stored in the database to the associated textual description. This file contains the translations for all coded fields in the database. A PROC file contains a saved ARPS query or portion of a query. Each PROC is highly encoded (i.e., the fields used in the selection and those to be displayed are represented by the offset from the beginning of the database record and length) and must be generated by hand.
- Typical Routine Reports. Much of the reporting done from the ASMIS system is done as routine reports. This section lists a selection of these reports.
- A figure which shows the data flows within the database. If necessary a second figure shows the daily processing associated with the database.

3.2.1. Ground Accident Database

This database contains data related to military ground accidents. Figure 3.3 shows the data flows associated with this database. Figure 3.4 depicts the daily processing.

This database consists of four data files. The accident data is stored in three files. The 285 file contains coded data plus short textual descriptions. The 285 narrative file is the narrative for the accident. The 285-1 is provided by professional accident investigators for only the most severe accidents and a random sample of other accidents. The exposure file contains the manhours worked and miles driven, and is used to calculate accident rates. Table 3.1 describes the size and growth rate of these files.

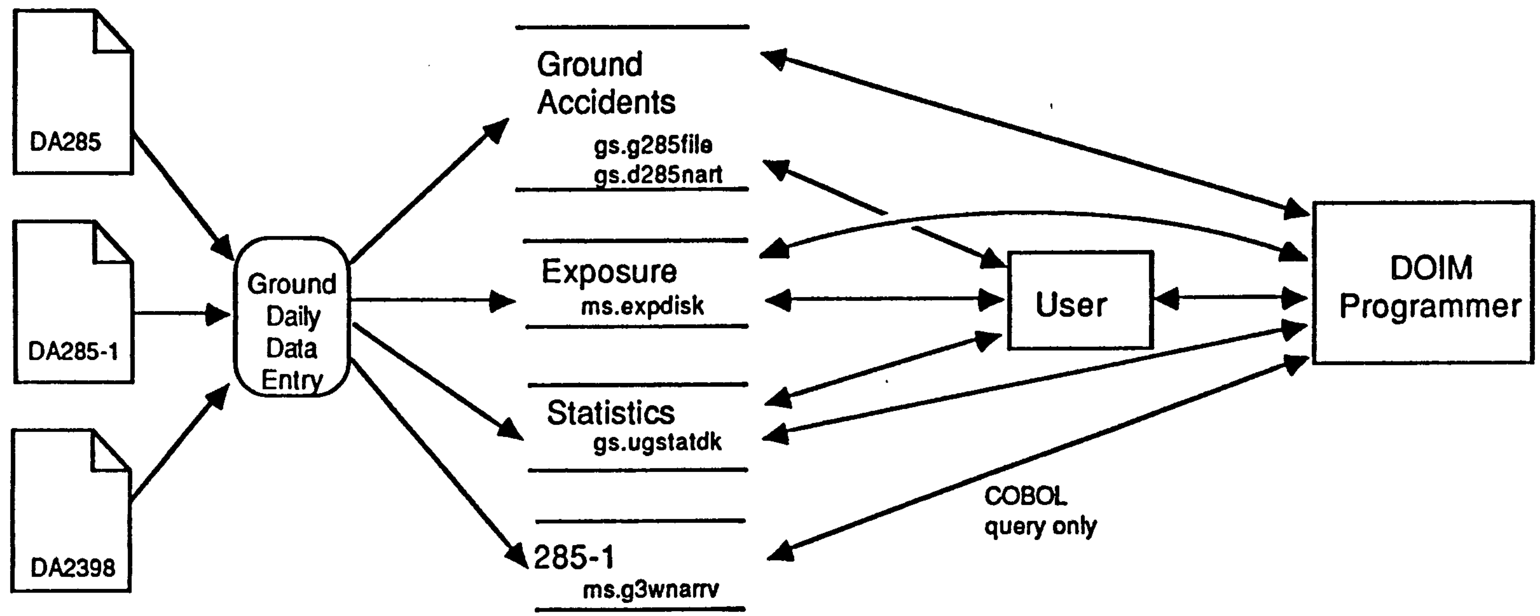
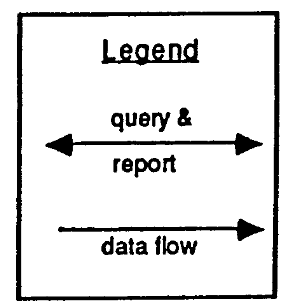


FIGURE 3.3 Ground Database



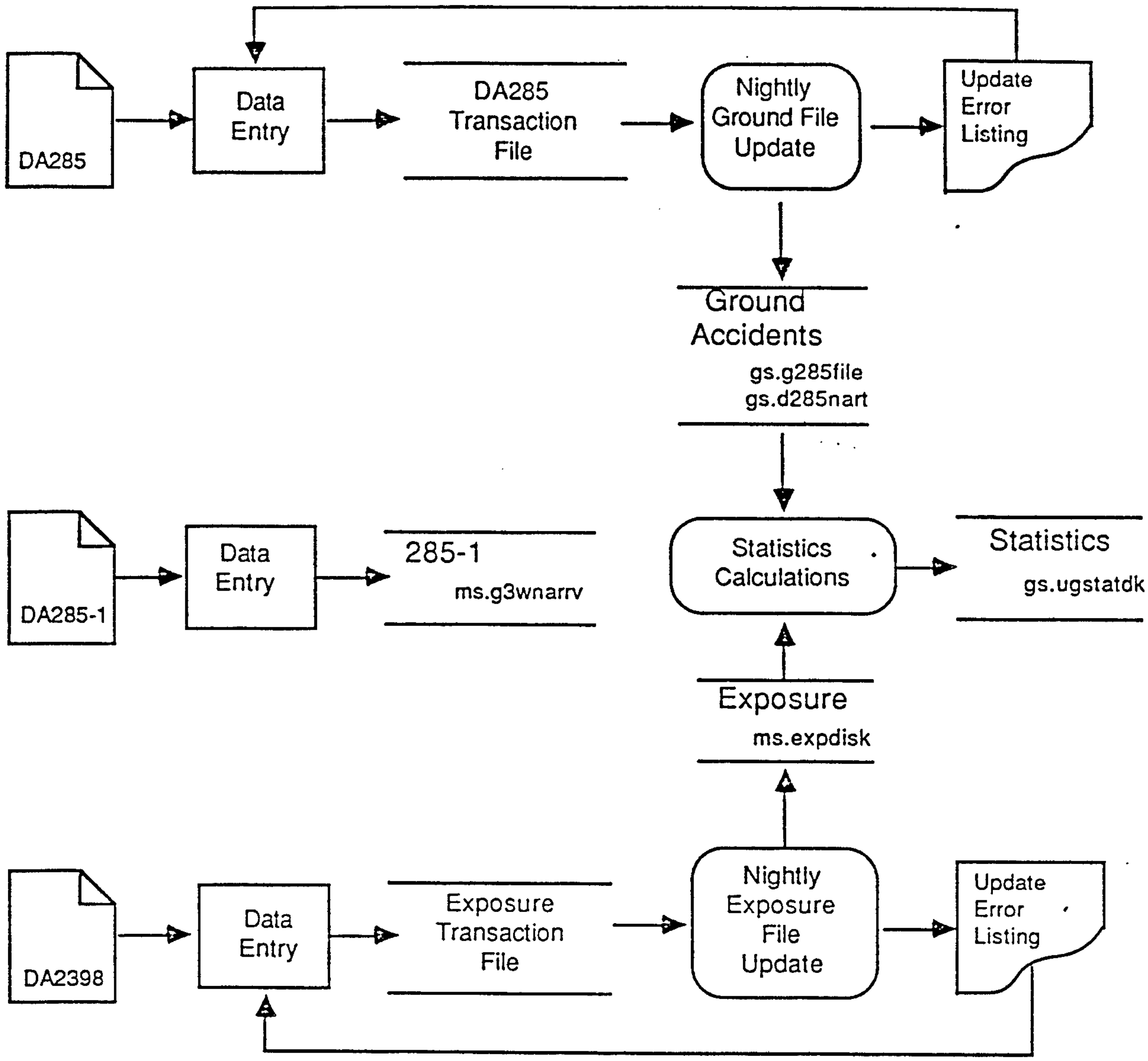


FIGURE 3.4 Ground Daily Data Entry

TABLE 3.1. Data Files Used, Number of Current and Historical Records, and Growth Rate for the Ground Accident Database

Conceptual Name	On-line File Name	On-line No. Recs (81-present)	Off-line No. Recs (74 - 81)	Growth Rate Recs/Year
285 file	gs.g285file	143k	120k	20k
285 narrative	gs.d285nart	660k	550k(a)	92k
285-1	ms.g3wnarrv	10.7k	none	
Exposure	ms.expdisk	27k	none	

(a) The number of historical narrative records was estimated (based on the number of historical 285 records and number of narrative records per on-line 285 record). The growth rate was estimated (based on the number of narrative records per on-line 285 record).

A number of auxiliary files provide information used in accessing the ground database and in translating the codes used into English phases. These files are shown in Table 3.2.

Approximately 70 routine reports are associated with the ground database. Table 3.3 is a partial list of reports that are routinely generated.

TABLE 3.2. Auxiliary Files Associated With The Ground Accident Database

Conceptual Name	File Name	No. Recs
Ground Data Dictionary	gs.n285dbc	228
Exposure Data Dictionary	ms.expdbc	28
Code book	gs.codefile	9.6k
PROC File	gs.gtpproc	1522

TABLE 3.3. Typical Reports Generated From The Ground Accident Database

Nightly:	Update Errors
Weekly:	Fatality Listing Class A Summary
Monthly:	Fatality Listing List of Late Reports for FORSCOM Military Parachute Report Log of Accidents by States for National Guard Sport Injuries (USASC)
Quarterly:	Command and Installation Reports National Guard Accident Experience Report EUR Accident Injury Data for USAEUR Tank Weapon Accidents Military Parachute Report National Guard Accident Exposure Selected Pieces of Materiel Report

3.2.2. Aviation Accident Database

This database contains data related to military aviation accidents. Figure 3.5 shows the data flows associated with this database. Figure 3.6 depicts the daily processing.

This database consists of eight files. The computerized information for each accident report is stored in six files; the basic, miscellaneous, personnel, impact, narrative and three W files. Since 1983, the narrative account of the investigation reported on the DA2397 form has not been computerized. The pre-1983 narratives are still available on-line.

One additional file, the cross reference, allows the use of a nine character accident identifier. It translations from model, type and serial for the plane plus date of accident into a six digit accident identifier, two digit sequence number and a single digit plane number (to differentiate planes in accidents involving more than one). This reduces the unique key for each record from approximately 24 to 9 characters and is done to save disk storage space.

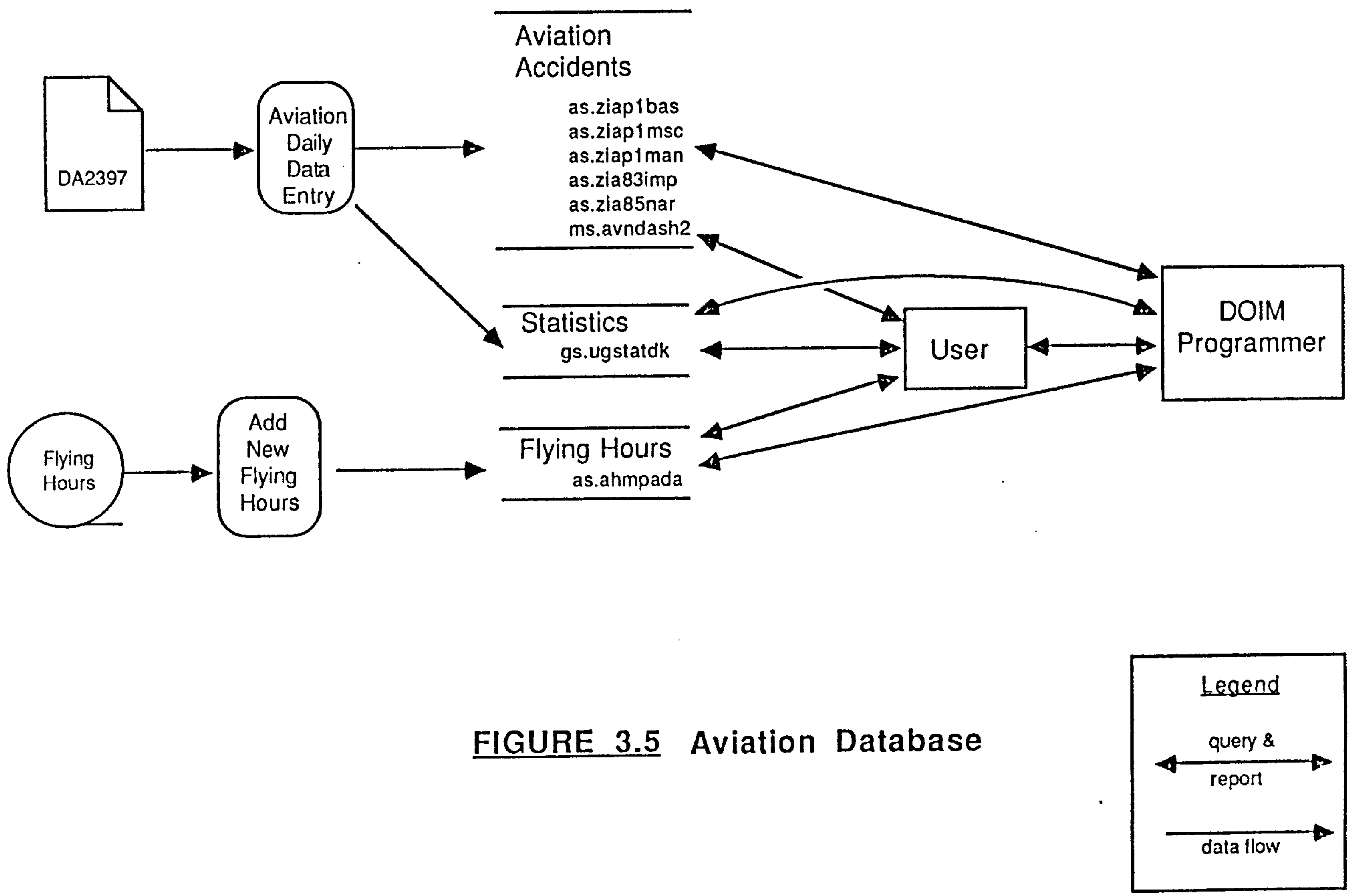


FIGURE 3.5 Aviation Database

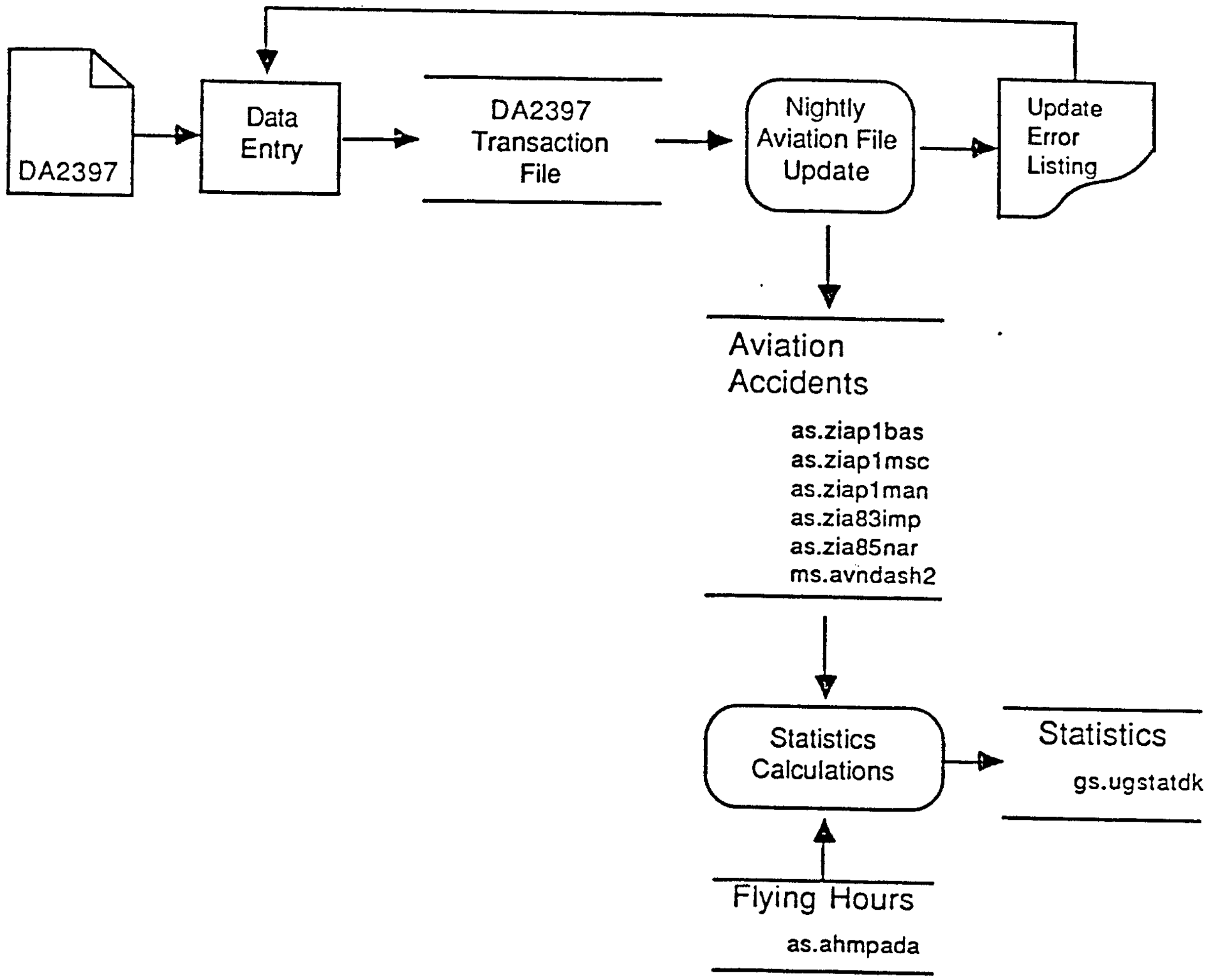


FIGURE 3.6 Aviation Daily Data Entry

The flying hours file contains the number of flying hours and landings used to calculate accident rates. Table 3.4 describes the size and growth rate of these eight files.

A number of auxiliary files provide information used in accessing the ground database and in translating the codes used into English phases. These files are shown in Table 3.5.

Approximately 80 routine reports are associated with the aviation database. Table 3.6 is a partial list of reports that are routinely generated.

TABLE 3.4. Data Files Used, Number of Current and Historical Records, and Growth Rate for the Aviation Accident Database

Conceptual Name	On-line File Name	On-line No. Recs (72-present)	Off-line No. Recs[a]	Growth Rate Recs/Year
Basic	as.ziap1bas	66k	31k	4.8k
Misc	as.ziap1msc	36k	49k	2.6k
Personnel	as.ziap1man	52k		3.8k
Impact	as.zia83imp	1.5k		0.1k
Narrative	as.zia85nar	91k		6.6k
3 W file	ms.avndash2	1k		
Flying Hours	as.ahmpada	9.5k	none	
Cross Ref.	as.xref	193K		

(a) Data for all aviation accidents recorded on the current form are on-line. The historical files are from a previous revision of the form. The coding conventions are not completely compatible. They are not available under ARPS.

The growth of each file was estimated from estimated number of accidents per year (4800) and ratio of the number of records in the file of interest and the basic file. For example, growth of the personnel file is 52000/66000 * 4800 = 3.8k.

TABLE 3.5. Auxiliary Files Associated with the Aviation Accident Database

Conceptual Name	File Name	No. Recs
DA2397 Data Dictionary	as.avndbc	1107
Flying Hours Data Dictionary	ms.fltdbc	17
Code book	as.codefile	8k
PROC File	as.atpproc	462

TABLE 3.6. Typical Reports Generated From the Aviation Accident Database

Nightly:	Update Errors
Weekly:	Mishap Summary
Monthly:	Mishap Summary by Type, Model, Series (for some users) National Stock Number Report (for some users) POL Problems
Quarterly:	Case Prints Sent to Aircraft Vendors Mishap Summary by Type, Model, Series (for some users) National Stock Number Report (for some users) Aircraft Mishaps by Fiscal Year
Annual:	Army Weather Related Mishaps (to Ft. Rucker)

3.2.3. FECA Accident Database

This database contains data related to civilian accident claims. Figure 3.7 shows the data flows associated with this database and the periodic processing (monthly or quarterly) that is done.

Data on civilian accidents comes in two kinds. The FECA monthly Table II tape contains information on accident claims filed. The FECA quarterly chargeback tape contains information on claims paid. Because of the difference between a fiscal year (October through September) and a calendar year (January through December), the quarterly chargeback data is not available under ARPS. This data is available on the system and quarterly reports are sent to each Army installation. Table 3.7 describes the size and growth rate of these files. Table 3.8 describes the auxiliary files associated with this database.

TABLE 3.7. Data Files Used, Number of Current and Historical Records, and Growth Rate for the FECA Accident Database

Conceptual Name	On-line File Name	On-line No. Recs (85-present)	Off-line No. Recs (83-85)	Growth Rate Recs/Year
Feca Table II	msowcptab2	60k	?	
Feca Quarterly	msfecaqtrs	112k	?	

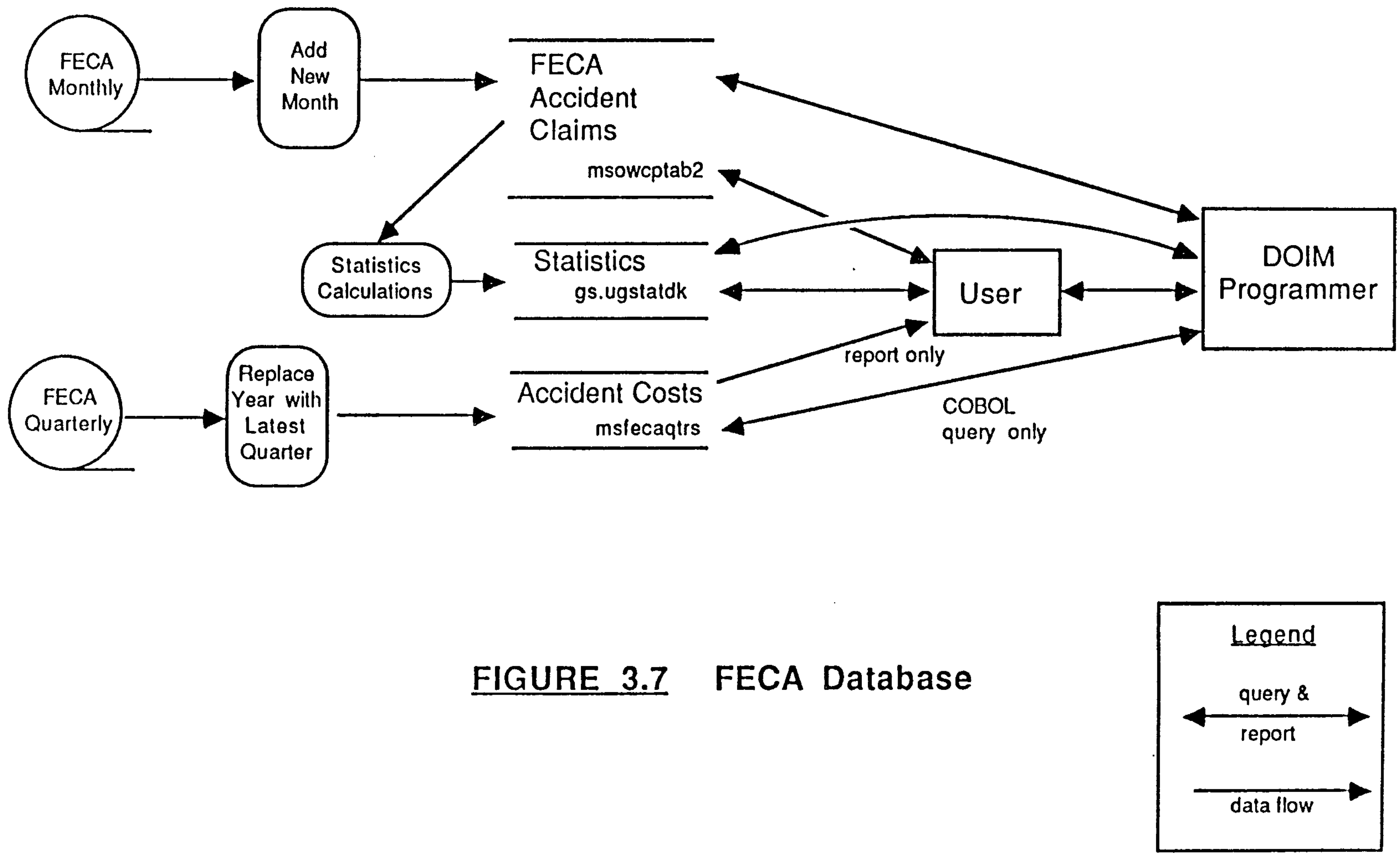


FIGURE 3.7 FECA Database

TABLE 3.8. Auxiliary Files Associated with the FECA Accident Database

Conceptual Name	File Name	No. Recs
Table II Data Dictionary	msowcpdbc	165
PROC File	msowcpproc	28

Routine reporting is very limited. Table 3.9 is a partial list of reports that are routinely generated.

TABLE 3.9. Typical Reports Generated from the FECA Accident Database

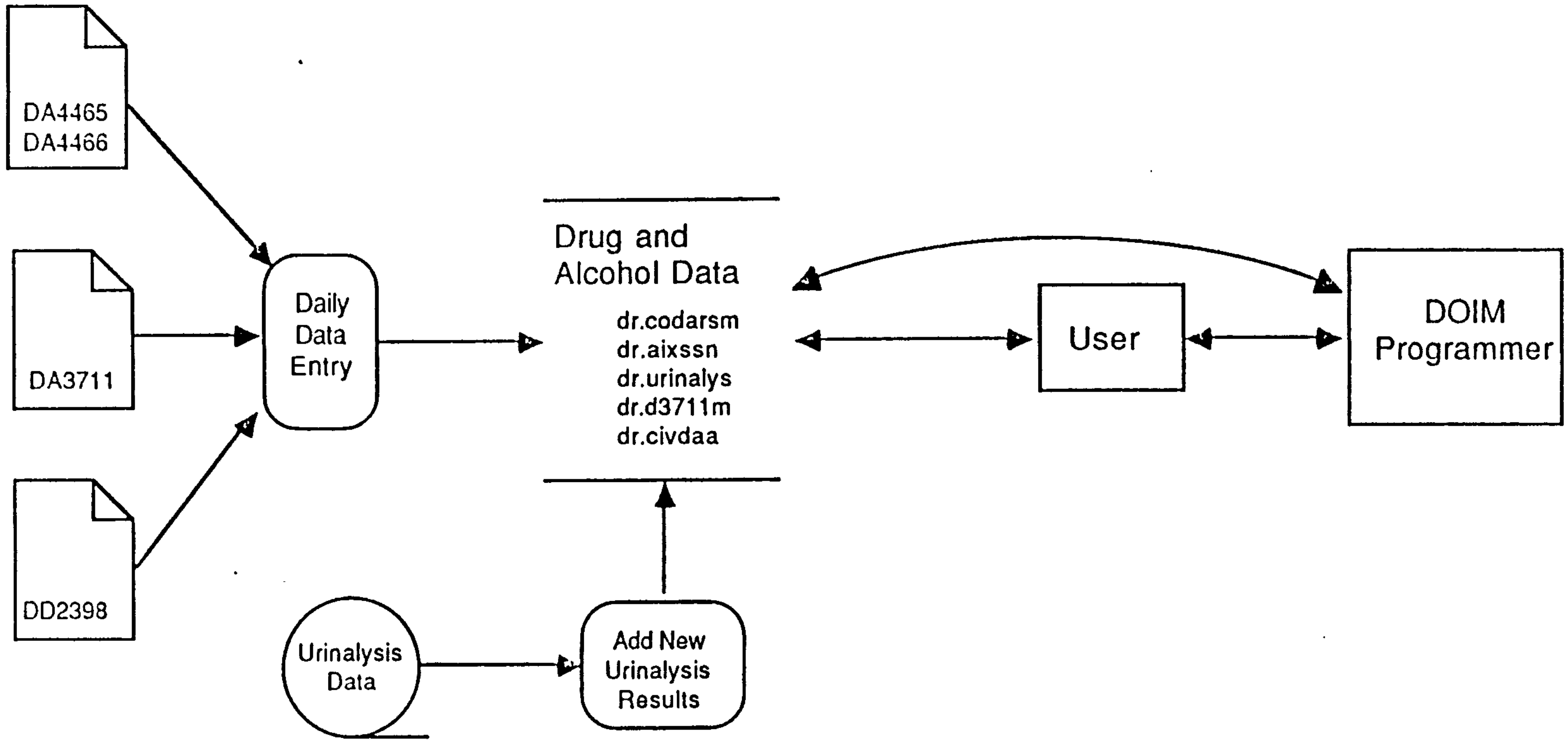
Monthly:	Monthly Reports
Quarterly:	Quarterly Reports

3.2.4. Drug and Alcohol Database

This database contains data related to the use of drugs and alcohol by Army personnel and civilian employees of the Army. It includes information on individual clients as well as on the facilities available for intervention. Figure 3.8 shows the data flows associated with this database. Figure 3.9 depicts the daily processing.

The data is stored in four files. The Client Oriented Drug and Alcohol Reporting System (CODARS) file contains information on the enrollment and progress reports for individual military clients. The Urinalysis file contains test results for individual military clients. The DA3711-R describes the Alcohol and Drug Abuse Prevention and Control Program (ADAPCP) facility, its staffing, services provided, number of pending cases and testing done. The DD2398 file summarizes the civilian client information for an ADAPCP. Table 3.10 describes the size and growth rate of these files.

Table 3.11 lists the auxiliary files associated with the Drug and Alcohol database. Table 3.12 is a partial list of the reports that are routinely produced.



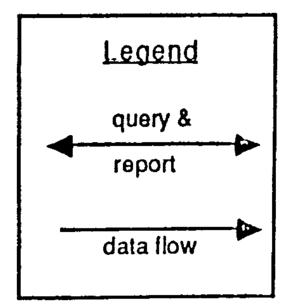


FIGURE 3.8 Drug and Alcohol Database

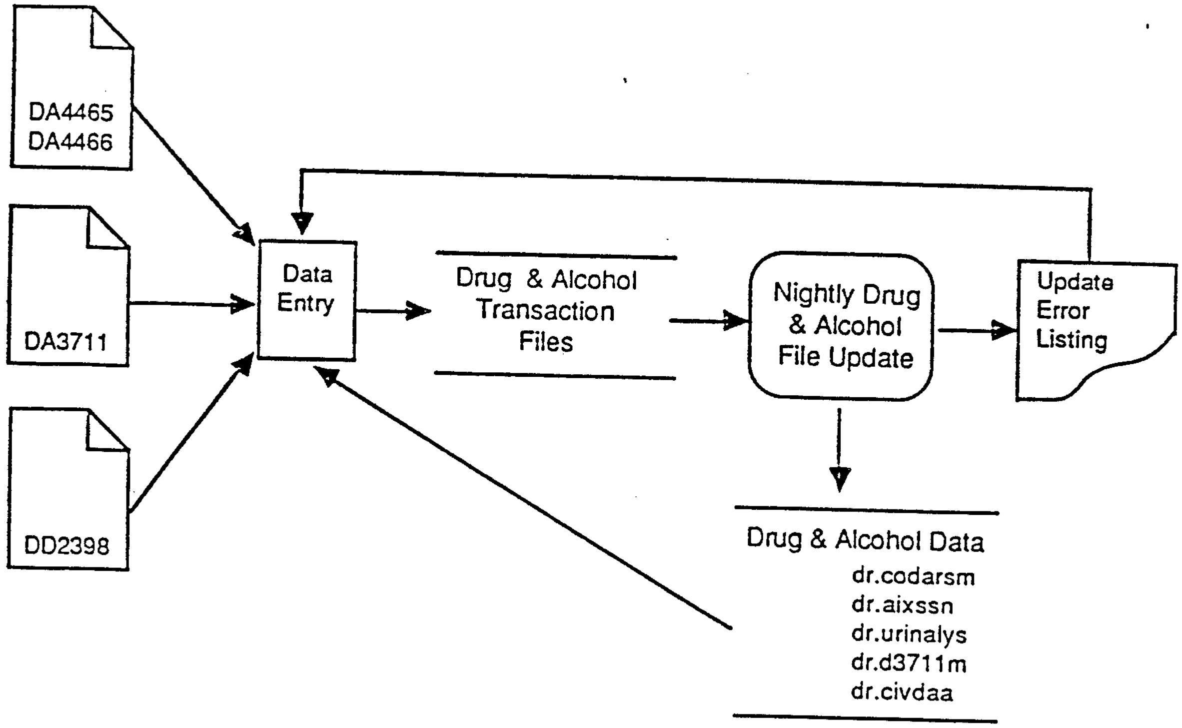


FIGURE 3.9 Drug and Alcohol Daily Data Entry

TABLE 3.10. Data Files Used, Number of Current and Historical Records, and Growth Rate for the Drug and Alcohol Database

Conceptual Name	On-line File Name	On-line No. Recs (84-present)	Off-line No. Recs (81 - 83)	Growth Rate Recs/Year
CODARS	dr.codarsm	286K	85k	60k
Index to CODARS	dr.aixssn	244K	?	
Urinalysis	dr.urinalys	128K	none	
DA3711-R	dr.d3711m	5.3k	none	
DD2398 (a)	dr.civdaa	1	none	

(a) This type of data is not currently being computerized.

TABLE 3.11. Auxiliary files associated with the drug and alcohol database

Conceptual Name	File Name	No. Recs
CODARS Data Dictionary	ms.dalcndbc	117
Urinalysis Data Dictionary	da.urindbc	17
DA3711 Data Dictionary	da.dbc3711	220
Code book	dr.codebook	2.5k
Urinalysis PROCS	da.urinproc	11
3711-R PROCS	da.proc3711	0

TABLE 3.12. Typical Reports Generated from the Drug and Alcohol Database

Nightly:	Update Errors
Quarterly:	Info Papers
Semiannual:	DOD Report
Annual:	Info Papers

3.2.5. Files Common to Multiple Databases

There are two files which are used by multiple databases. Table 3.13 describes the sizes of these files. When an accident is reported, the six current UICs (UIC1, the major command (MACOM), to UIC6, the unit) are reported. Over time, the Army has reorganized and thus these UIC designations assign the accident to an obsolete group. The UIC Translation provides a way to look at the unit, UIC6 and place this unit within the correct hierarchy for the current Army organization. This information is used in the calculation of safety goal numbers (see Section 3.3.4 for details).

The Statistics File is a collection of reports which are precomputed each night. This file contains information summarized from the Ground, Aviation and FECA databases. The purpose of this file is to provide answers to often-asked questions and thus reduce the workload of the ARPS query processor. Section 3.3.3 contains more details on this process. Figures 3.4, 3.6, and 3.7 show the data flows surrounding the generation of this file.

TABLE 3.13. Data Files Used, Number of Current and Historical Records, and Growth Rate for Information Associated with Multiple Databases

Conceptual Name	On-line File Name	On-line No. Recs	Off-line No. Recs	Growth Rate Recs/Year
UIC Translation	gs.uicmastr	8.1k	none	
Statistics File	gs.ugstatdk	143k	none	none

3.3. ROUTINE PROCESSING

In addition to the ad hoc query capability, the USASC provides an extensive list of routine services. Computer operators are available both first and second shift to facilitate these operations. The daily operations are:

- Entry of data from paper forms is done during first shift. The results of this entry (both new records and corrections) are held in transaction files which are merged with the master files during the second shift.
- Disk backups are done during second shift.
- Generation of the Statistics file is done during second shift. It is done after the processing of the data entry transaction files, so that the statistics available will reflect the current database.
- During second shift, the historical portion of the databases are reloaded onto a scratch disk. Query jobs requiring the historical data are then

run during third shift (no operator is available). These historical files are removed from the scratch disk at the start of the first shift.

- During third shift, all routine reports are run. These jobs are scheduled and submitted by DOIM personnel.

3.3.1. Data Entry

3.3.1.1. Data from Paper Copy

Data from the DA285, DA285-1, DA2397 and DA2398 forms is entered by the Accident Records Management Division of USASC. Currently submission of the DA285 form via electronic transfer (either using a modem or by sending a diskette) is being tested. Data from the DA4465, DA4466, DA3711-R and DD2398 forms is entered by USADAOA in Washington, D.C. Each of these types of data, except the DA285-1, is processed in two stages. During the day, new data, corrections and deletions are entered and held in a transaction file. At night, the transaction file is merged into the master file(s). The DA285-1 data is entered directly into the master file.

The validation of the data entered is also done in two stages. During the actual entry of the data record (during the day), each coded field is checked against the list of legal codes (stored in the code book). During the nightly transaction processing, cross field validation is done and an error report is returned to the data entry personnel the following morning.

In general, updating is done using the transaction file process described above. The ability to make corrections directly into the master file for the database for a limited number of fields is available. This is rarely used.

At times, the processing of corrections to the various databases is halted. This is an administrative decision and is typically done for major report generation, for example, the IPR or the annual Army Safety Report. Freezing database corrections provides a means of maintaining a consistent set of data. The addition of new data is not halted, rather the selection criteria for each data request must include the SDATE (date case established) to control which accident records are retrieved and used in the report.

3.3.1.2. Data from Magnetic Tapes

Adding data from tapes to the various ASMIS databases is a simple process. A COBOL job adds the new data to the master file while doing a few simple transformations. For the data from AVSCOM, the Urinalysis laboratories and FECA (monthly), data is added to the master file. For the FECA quarterly tapes, the data is cumulative for the Department of Labor fiscal year (July to June). Second, third and fourth quarter tapes replace data for the year. The first quarter tape creates records for the new year.

3.3.2. Reporting

The Data Processing Division is responsible for scheduling and running a large number of routinely generated reports. There are approximately 80 reports associated with the aviation database, 70 associated with the ground database, 2 associated with the FECA database and 15 miscellaneous reports. The miscellaneous group contain some reports which use multiple databases (e.g., the Precommand course stats) and some which print information from various support files (e.g., the UIC lists and printing of code books).

Scheduling of these reports is carefully coordinated with the data availability. For instance, delay in processing the flying hours tape might result in delaying reports which use this data to calculate accident rates.

Because of the dynamic nature of corrections to the databases, the reports for the same destination are run on the same night to guarantee the consistency of numbers from report to report.

3.3.3. Statistics Generation

Nightly, after the transaction file processing for all databases is completed, the statistics data is regenerated. This is a large collection of preformatted reports for aviation, ground and FECA. The stated purpose of this file is to provide answers for many repetitive ad hoc queries. The pregeneration of these answers saves processing time during the day and thus reduces the time required to respond to these queries. Figure 3.10 shows the data flows for this process. Because the statistics file contains data from

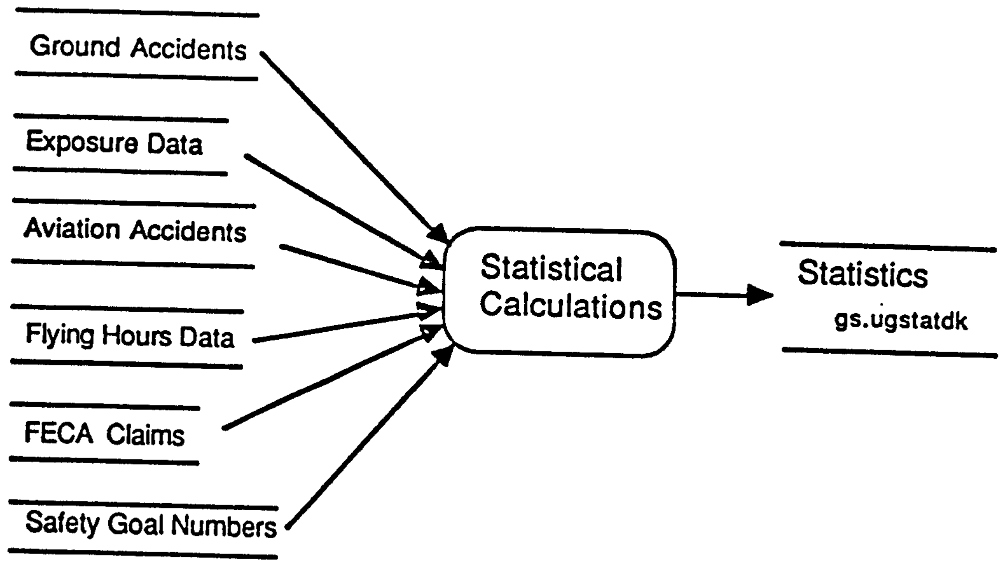


FIGURE 3.10 Generation of Statistics file for use in ARPS

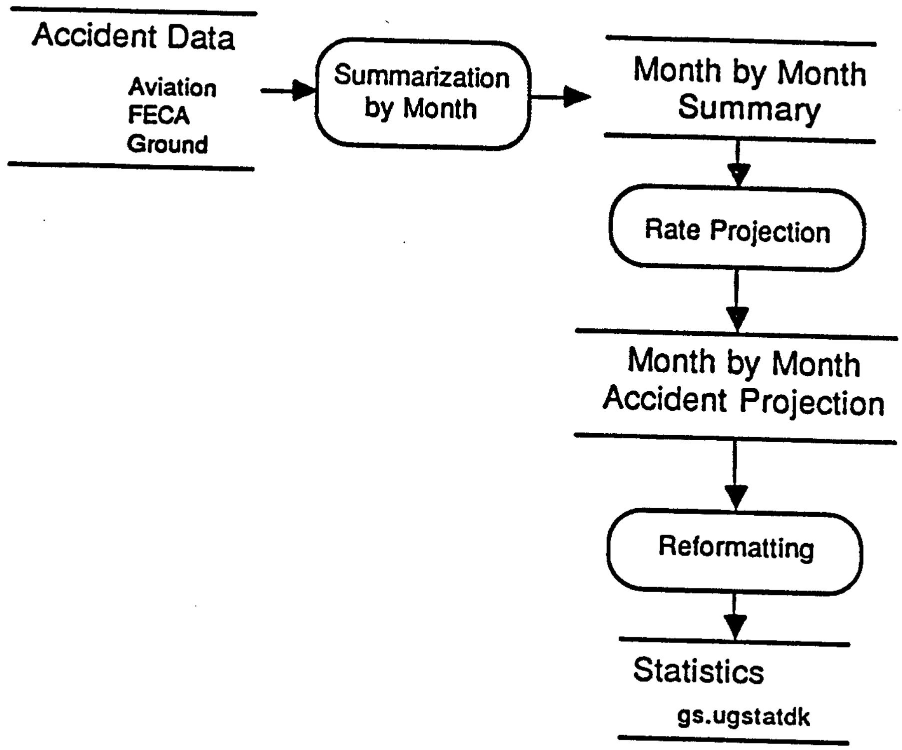


FIGURE 3.11 Creation of Safety Goal Numbers

the aviation, ground and FECA databases, the file is shown on the figures for each database (see Section 3.2.1, 3.2.2 and 3.2.3 for more details).

3.3.4. Safety Goal Generation

Monthly the safety goal numbers are generated and incorporated into the statistics file where they are accessible under the ARPS statistics option. Quarterly the goal numbers are generated and incorporated into a report to be distributed. Figure 3.11 shows the process used. The accidents, projected accidents and goals are reported as numbers not as rates. Before the goal and accident numbers are generated, each accident is moved from the UIC structure applicable when the accident was reported to the current UIC structure. The information to permit this translation is in the UIC translation file.

3.4. ACCESS METHODS USED TO RESPOND TO USER REQUESTS

User requests come in two different forms. Many users satisfy their requirements by calling or writing USASC and having a Data Processing Division staff member answer their query. The DOIM staff member uses the ARPS statistics reporting option, the ARPS ad hoc query facility, SAS or a COBOL program to satisfy the request. Other users log onto the computer directly and pose their own queries. These users can access either the statistics reporting option or the ad hoc query facility within ARPS. They can not use COBOL programs.

3.4.1. ARPS

ARPS provides two basic functions. It provides the display of portion(s) of the statistics or safety goal information and a general purpose query processor with a limited reporting capability. The display of the statistics and safety goal information is accomplished by asking questions to isolate the report to be displayed and then simply displaying it. The ARPS query processor provides three types of output: a case list (the specified fields for each selected case are printed on a single line), a two- or three-dimensional matrix or a caseprint (all fields for the case are printed).

Case listings display a maximum of 132 characters of information for each case (exclusive of narrative, which is listed on multiple lines below the other displayed fields). By default, this listing is in case order. It can be sorted, but the sort has limitations. You are allowed to sort an unlimited number of records, but the sort key is limited to 30 characters. Any query which involves a sort is reassigned to the delayed processing mode (the job runs as a batch job).

The matrix is a two-dimensional display of values of Field A against values of Field B. It is limited also. The matrix is limited to 500 rows by six columns (larger matrices are handled by SAS). This severely limits the field to be used as the horizontal variable. You can specify the values to be displayed by either listing the individual values or creating ranges of values. For fields with many values, multiple matrices would be necessary to get all of the individual values for the horizontal field reported.

A three-dimensional matrix can also be generated by specifying the field name for the third dimension as a "field to be displayed," followed by the keyword "BREAK" before specifying the matrix in the normal manner.

ARPS provides three modes of query processing. A query can be submitted for interactive response (known as terminal), for delayed terminal display (you get your response tomorrow as a file you can view at the terminal) or mailed (your output is printed and mailed to you).

ARPS provides access to both the on-line and off-line components of the databases. Access to on-line data can be accomplished interactively at the terminal. For user convenience, queries for on-line data can also be created in delayed or mailed mode. Access to off-line data is accomplished by creating of a delayed or mailed job. These jobs are then run overnight.

Routinely, these off-line records are copied to a scratch disk by the second shift operators and all queries needing this data are run during third shift. The first shift operators clear this data off the scratch disk.

3.4.2. Non-ARPS

Most non-ARPS access to the databases is done using COBOL programs. This access is used to generate routine reports, to access files which are not known to the ARPS query processor (e.g., 285-1 and the FECA quarterly files), and to answer user queries which require more complex reports than ARPS is capable of generating. More complex reports require longer sort keys, calculations, special printing (e.g., inclusion of dashes, print to imitate original input form), and decoding of fields.

Another way to process the data is to use ARPS to select a subset of data and write the fields for each case to a disk file. SAS is then used to produce the desired output. One such use is to produce a matrix when the field on the vertical axis of the matrix has more than 500 values.

3.5. DATA ANALYSIS FACILITIES AVAILABLE TO USERS

The user has three methods of manipulating the data from his request. He can use the case list or matrix generated by ARPS directly, transfer the ARPS output into SAS or download it onto his personal computer (PC). To transfer into SAS or download to a PC, a small number of PROCs exist which list the data without headings. Each PROC contains a specific list of fields to be displayed and is still limited by the 132-character limitation for the case list.

3.6. USER DOCUMENTATION

Documentation for the ARPS system is available in three manuals: "Ground User's Guide," "Aviation User's Guide," and "Feca User's Guide." In addition, all new users of the ASMIS system get the "Introduction to Computer Literacy."

3.7. DATA SECURITY MECHANISMS

3.7.1. Backup and Archiving of Data

The system manager provides a full system backup once a week. This is accomplished by backup of several actuator a night on a rotating basis. Nightly incremental backups of the other actuators are also done. Thus, reconstruction of the contents of an actuator can be accomplished by reloading the most recent full backup and adding to it from the incrementals for that actuator. Multiple version are kept.

The programmer responsible for a particular database or file within a database is responsible for backups related to major changes in the file. Multiple versions of each file are kept.

Weekly, database backup tapes are moved from the USASC to another building to provide backup in case of destruction of the building. The system backup tapes are not now being moved to another building.

3.7.2. File Protection

The programmer responsible for a particular database or file within a database is responsible for establishing and maintaining file protection passwords.

3.7.3. ARPS Security Features

The ARPS program limits access to certain fields. The data dictionary has an access level code which describes the type of user who can have access to that field. The classes of users are:

- General User. No access to sensitive information (name, social security number, or for aviation, review board members) and limited access to the narrative.
- Limited. No access to sensitive information (name, social security number or for aviation, review board members), but access to more narrative than general users.
- Safety Center. No limitations; access to all fields.

3.8. COMPUTER SYSTEM AND SOFTWARE

3.8.1. Computer System

The ASMIS system is currently installed on an IBM 4381 computer system with 24 MB of memory running the MVS/SP operating system, Version 1.3.5. A diagram of the current configuration of the ASMIS computer hardware is shown in Figure 3.12. A list of the peripherals attached to the IBM 4381 with a description of how they are used is provided below:

- One IBM 3880 disk controller with three IBM 3380 disk drives, each having 2.5 gigabytes of storage for a total capacity of 7.5 gigabytes of disk storage.
- One IBM 3480 magnetic tape controller with three 3480 cartridge tape units. These tape units are very reliable and the cartridge tapes themselves are much more compact in size than the older reel-to-reel magnetic tapes. The tape units are used to make backups and store historical data.
- One IBM 3430 magnetic tape controller/unit with one IBM 3430 reel-to-reel magnetic tape unit. These tape units are primarily used to exchange data tapes with external sources.
- One IBM 2821 unit controller with two 1403 printers. These printers are used to provide hard copy output of data sets and reports to users of ARPS, including off-site users who receive the output by mail. The 2821 controller and 1403 printers are scheduled to be replaced with a Xerox 4060 ion deposition page printer (similar to a laser printer) later this year.
- One IBM 3725 communication controller running the Network Control Program (NCP) with the Network Terminal Option (NTO). (For more information about ACF/NCP and NTO refer to the description under computer software). Eight modems are connected to the 3725 controller to allow off-site access. Two of the modems are designated as 300 baud rate lines and six are designated as 1200 baud rate lines.
- One IBM 3708 Network Converter. This converter will not be available until later this year. When it is operable, the modems will be attached to the IBM 3708, which will be physically attached to the 3725 communications controller. Unlike the IBM 3725, the IBM 3708 allows modems to change baud rates as needed.

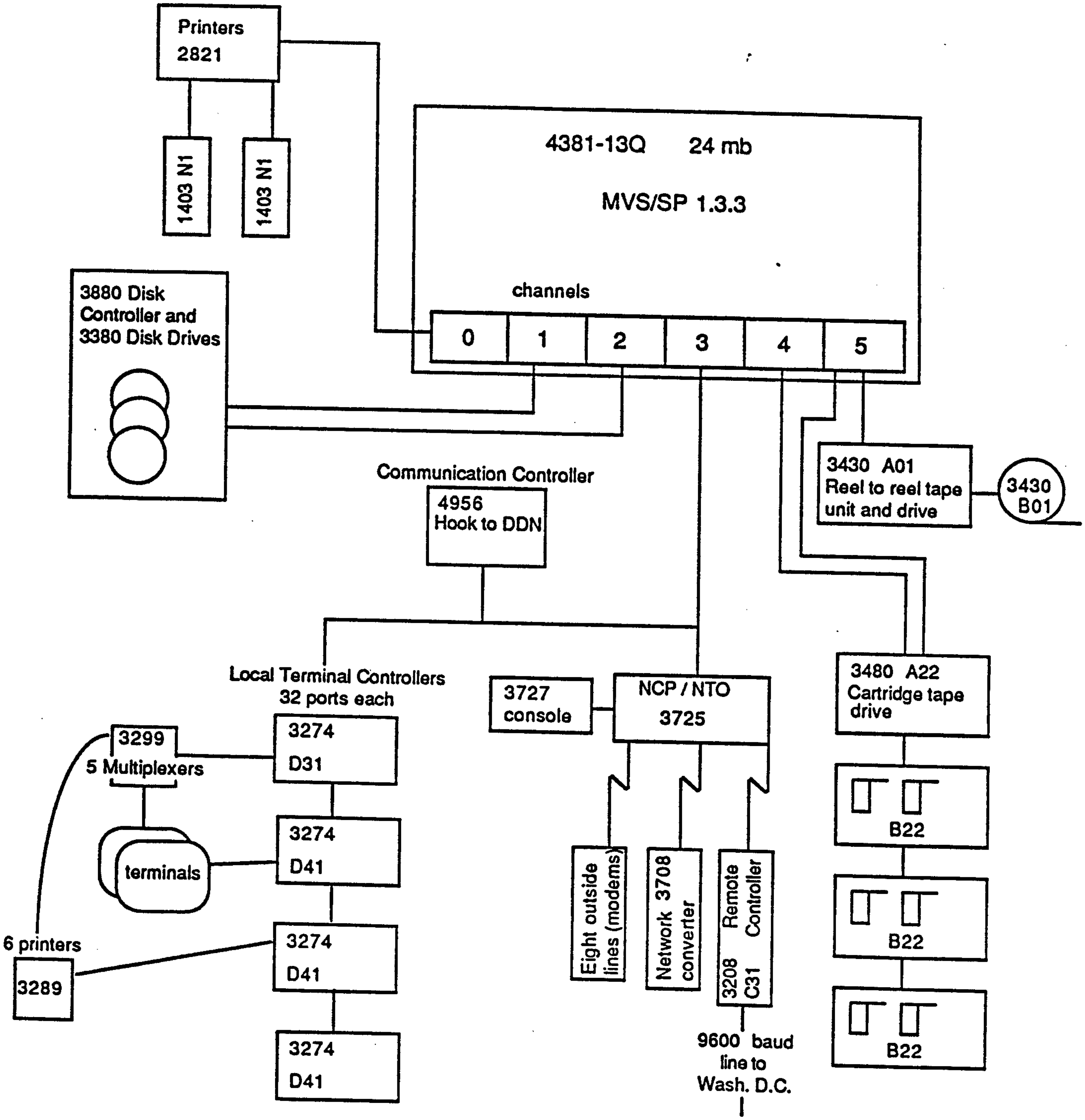


FIGURE 3.12 USASC Computer Hardware Configuration

- One IBM 3274 Remote Controller, which is attached to the 3725 communication controller. This device is used to provide an SNA 9600 baud rate line to Washington D.C. for Drug and Alcohol data update and retrieval.
- Four IBM 3274 control units. Each unit allows up to 32 devices to be directly connected for a total of 128 devices.
- Five IBM 3299 Terminal Multiplexers. Each multiplexer is connected by coaxial cable to a single port on one of the IBM 3274 control units, and each multiplexer has eight ports to connect terminals and printers. There are over 100 terminals connected through the IBM 3274 control units, either directly or through the IBM 3299 Terminal Multiplexers.
- Six IBM 3287 Matrix Printers. These printers are used for on-site Army Safety Center use. They can be attached to either the IBM 3274 control units or the IBM 3299 terminal multiplexers. One printer is directly attached to the IBM 4381 computer system.
- One IBM 4956 Series-1 Communications Controller. This controller is used to provide access to the Department of Defense Network (DDN).

3.8.2. Computer Software

The following computer software is available on the ASMIS IBM 4381:

IBM products:

- MVS/SP (Multiple Virtual Storage/System Product) operating system.
- EREP (Error Recording Editing and Printing) is a reporting program for hardware and software errors detected by the operating system.
- ACF/VTAM (Advanced Communications Function/Virtual Telecommunications Access Method) controls the network that connects terminals to the mainframe and the programmatic interface from applications (called VTAM applications) to the terminals.
- ACF/NCP (Advanced Communications Function/Network Control Program) resides in the 3725 Communication Controller and provides the physical management of the communication network. ACF/NCP controls attached lines

and terminals, performs error recovery, and routes data through the network. ACF/NCP communicates with the host through ACF/VTAM.

- ACF/SSP (Advanced Communications Function/System Support Program), a set of programs used by IBM system engineers during software installation.
- NTO (Network Terminal Option) extends the capabilities of ACF/NCP to allow access to certain non-SNA (System Network Architecture) terminals, (such as ASCII), through the record mode application program interface in ACF/VTAM. NTO preserves the non-SNA data stream, which minimizes changes to existing application programs.
- RMF (Resource Management Facility) collects data and provides reports on the performance and operation of the MVS/SP operating system.
- DFP (Data Facility Product) provides data management, device support, utility functions, and user and system catalog support.
- ISPF (Interactive Structured Productivity Facility) is a menu-oriented program which operates under TSO and provides dialogue management services to users of IBM 3270 Display Terminals. Dialogue management services may be used by an application developer to produce interactive applications in the form of menu-driven dialogues.
- ISPF/PDF (Interactive Structured Productivity Facility/Program Development Facility) provides menu-driven design and library management facilities oriented toward improving programmer productivity.
- CICS (Customer Information Control System) provides two general sets of functions:
  - A set of callable functions that greatly reduces the effort otherwise needed for terminal-oriented transaction programming.

- A set of functions to manage the network, terminals, storage, transaction recovery, etc. Some of the key resource and performance functions that CICS provides are:
  - -- Dynamic multitasking to provide greater throughput
  - -- Quasi-reenterable programming to optimize storage usage
  - -- Intersystem communications

- SMP/E (System Modification Program/Extended) is a product which is used to maintain MVS system components. SMP/E records all changes to a component and ensures that the component is at the proper level to receive new maintenance.
- COBOL (Common Business-Oriented Language) compiler and library. This is the primary language used for application development by ASMIS.
- COBOL (Common Business-Oriented Language) debugger. This utility is used to assist in debugging application programs.
- TSO COBOL Prompter is an interactive interface for TSO COBOL users which creates JCL (Job Control Language) for the COBOL compiler.
- Host to PCXT transfer program provides PC's with 3270 emulation boards and the corresponding PCXT to Host transfer program to download and upload files to the IBM 4381 mainframe computer.
- DASD Migration Aid was used in the migration from the 3330 disks to the 3380 disk drives. This software is no longer being used.
- OS Assembler translates assembly language into machine language which runs on the MVS operating system. Assembly language is a low-level language and is used to produce efficient code.
- PL/1 Transient Library - this software is required by the statistical package SAS.
- GSAM - Global Storage Access Method. Allows multiple users to access the same VSAM files for update. This product is used in two or three TSO data update applications. These applications are in the process of

being transferred to the CICS environment. When this is completed, GSAM will no longer be used.

- TSO DSPRINT a print utility which allows TSO applications to print files on the IBM 3287 matrix printers.

Software on the 4956 Series-1 controller:

- DDN/MVS software allows communication to the Department of Defense Network.

Other Vendors:

- SIM 3278/VTAM is manufactured by Simware Inc. This program is a multiple session manager allowing users to run multiple CICS and TSO applications at one time. It also provides full screen support for ASCII terminals.
- SAS (Statistical Analysis System) is manufactured by SAS Institute Inc. SAS is an all-purpose data management and analysis tool.
- SAS/GRAPH provides an easy mechanism to produce charts, plots, three-dimensional grids and maps.
- SAS/FSP provides full-screen, interactive data entry, editing, and querying capabilities.
- SAS/AF provides an interactive, full-screen facility to develop applications.
- Panvalet and Panvalet TSO option are manufactured by Pansophics. Panvalet is a code management software package. It is used to maintain a library of source code, keep track of different versions, and control a multiple-user development environment.
- Easytrieve is manufactured by Pansophics. Easytrieve is a report writer which was used to generate reports for the Drug and Alcohol database. No new development is being done in Easytrieve.
- Syncsort is manufactured by Whitlow computer. Syncsort is a fast system sorting facility.

- Fast Dump/Restore is manufactured by Innovation. Fast Dump/Restore is a facility which allows the user to dump data from disk to tape and restore the tape back to disk. It performs reorganization of disks and alleviates disk fragmentation.
- Tape Management System is manufactured by Universal Computing. It keeps track of the owners of tape volumes.
- PROMAX is manufactured by Computer Associates. This software provides an easier way to develop user interfaces for application software running under CICS.
- TSO Superset Utility is manufactured by Applied Software. It provides a collection of data set utilities such as listing, merging, formatting and sorting files.

4.0 FUNCTIONAL REQUIREMENTS

This chapter will describe the functional requirements for four separate areas.

- the collection and computerization of the data
- the database system
- the user interface
- the computing environment that surrounds the ASMIS system.

Each section below will deal with one area of the functional specification for the ASMIS system. Each subsection includes a definition of the area to be discussed, a statement of the requirements for that area and a description of the features and problems of the current ASMIS system.

4.1. COLLECTION AND COMPUTERIZATION OF THE DATA

4.1.1. Data Acquisition and Preparation

Data acquisition is the process of collecting and organizing the facts about an accident and entering this information on the hardcopy forms. Army accidents occur world-wide and the associated data acquisition is also done on a world-wide basis. Data preparation is the process of converting the textual descriptions on the form to codes which are entered into the computerized database. Currently, this data preparation is done at the Safety Center.

4.1.1.1. Requirements

A common frame of reference for those acquiring and recording the data on the hardcopy form must be provided.

The description of the accident entered on the hardcopy form must be complete. Missing answers, imprecise information which can not be coded accurately (i.e., more than one code could be used) will lead to incomplete or inaccurate computerized records.

A common frame of reference for the coding activities of the data preparers must be provided.

Sensitive data, i.e., data which if released could cause harm to the individuals involved or to the Army, should be entered into data fields which are accessible to only authorized users (see Section 3.4.1 on data security). In particular, references to names should be eliminated from the narrative portion of the accident record. The name is a separate field in the record and is accessible to only authorized users. This information is available in two places, one accessible to only authorized users and the other accessible to anyone.

4.1.1.2. Features and Problems in the Existing System

The unit commander is responsible for completing the DA285 form. Providing a common frame of reference for the approximately 170,000 commanders is possible only as written instructions. Providing this framework for the DA2397 form is much easier. The form is completed by an accident investigator who has attended a course in completion of the DA2397 form.

The current data preparers are all within the Accident Records Management Division of the USASC. Procedures are in place to maintain a standard for the conversion of textual description to codes and for the addition of new codes.

The problem of missing or incomplete data is not easily remedied. Contacting the original recorder is a formidable task. The time between the accident and receipt of the form at the Safety Center is 1 to 360 days; the average is approximately 50 days. An answer obtained so long after the accident would not necessarily be accurate.

Currently, sensitive data is being entered into fields accessible by users with no special access privileges. The narrative for ground accidents often contains names. In some cases, names are replaced by a standard token, e.g., SM for service man, SM1, SM2, if multiple people are involved.

4.1.2. Data Entry

Data entry is the transfer of information from the hardcopy form to the computerized record. It includes the entry of the original data and all updates done to the record.

4.1.2.1. Requirements

Data entry, both the original entry and all updates should be validated. Validation includes:

- checking that the code entered is legal for the field
- comparison of two or more fields to verify correct relationship
- checking the whole form for completeness for the specified accident type.

The data entry person should be able to view the whole form when he/she is making updates.

Data consistency should be maintained.

Consistency between new data and old data should be maintained wherever possible. In some cases, consistency between old data and new data is not possible.

To improve data consistency, all fields which can be calculated from other fields should be calculated rather than entered. Calculated fields are those whose contents can be determined by applying some algorithm to one or more other fields in the record.

4.1.2.2. Features and Problems in the Existing System

The validation process is done in two steps. While the data entry person is actively entering the data, the validity of the code entered is checked against the list of valid codes. Any code not found is rejected and re-entry is required.

The second step of validation, checking one field against another, is done as part of the nightly transaction processing. The result of this validation is reported to the data entry personnel the following morning as a printed report. The data entry person is responsible for correcting the errors found. No computer enforcement of the correction is done.

Updates, except to the Drug and Alcohol Database records, are done without being able to view the existing record. The data entry person simply enters the correction and has no opportunity to review the other fields on the form.

New validations are being added to the nightly transaction run, but are not being applied to the whole database. For instance, in the ground database, total cost for the accident is not necessarily the sum of total damage cost and the personal injury costs for all individual involved. This validation is now in the nightly transaction procedure so that consistency is being maintained in the incoming data, but the check has not been run against the whole database to identify the inconsistencies in the old data.

4.2. THE DATABASE SYSTEM

4.2.1. Relationship Between the Database Redesign and the Data Forms

4.2.1.1. Requirements

The redesign of the ASMIS database will not change the basic data collection forms (e.g., DA285, DA2397) or the methodology for their use.

The Army Safety Center will make changes to the basic data collection forms. Currently, a draft of a proposed change to the DA285 form exists. The database must be able to adapt to these changes.

The current functionality of the ASMIS database will remain. The redesign of the database may change way the data is organized, but no data can be lost.

The current database contains a large amount of valuable data that must be preserved. This involves maintaining upward compatibility of the data as new data collection forms are designed and preservation of the existing data in spite of a move to the redesigned ASMIS system.

4.2.1.2. Features and Problems in the Existing System

Currently, there is no means to get more data directly from the data acquisition person (unit commander or accident investigator). The narrative portion of the form is used to derive additional information not explicitly on the form. For example the phrase, "No drugs or alcohol were involved," is often recorded as a derived variable by analysts. Encoding these additional data items in the narrative provides no means for enforcing the collection of this data. To obtain this data consistently would require a new form or a revision of an existing form.

The task of maintaining compatibility across a change in the data collection form is complex. As part of changing the ASMIS system to reflect the new DA285 (Oct 1980) the old data was translated into the format associated with the new form. This new format includes a section labelled "old system codes" which includes fields that could not be translated into the new format and fields that were dropped as part of the form revision.

4.2.2. Data Integration

Integration is the process of making the various parts of the database fit together. It includes making the connection between the parts obvious to the users and making these connections easy to incorporate in user queries.

4.2.2.1. Requirements

All data should be integrated. This includes integrating the data from the three current ASMIS databases (Ground, Aviation and FECA) into a comprehensive whole.

The historical data should be integrated with the current data.

4.2.2.2. Features and Problems in the Existing System

Currently ASMIS is an umbrella for four separate databases. Of these, three (Ground, Aviation and FECA) are closely related and the user should be able to access them together. Obtaining the number of accidents at an installation now requires three queries, one for each database.

Creating a ground database record which is a limited description of an aviation accident has been discussed. This duplication of data would create problems in keeping both copies of the aviation record correctly updated.

The file structure for the historical aviation data (the pre-1972 information) is different than that of the current aviation data (1972 to present). The pre-1972 data is currently not available under ARPS.

4.2.3. Type of Database Access

There are two basis types of database access. Read-only access implies no ability to insert, delete or change data. Write access allow insertion, deletion or updating of data.

Depending on the implementation methodology, access is limited to a select group of user or to a select group of programs. In general, database packages limit access by user. The current ASMIS system limits access by the program.

4.2.3.1. Requirements

Database access should be controlled on a user basis.

The number of users who have write access to the database should be limited. The majority of users should have read-only access.

4.2.3.2. Features and Problems in the Existing System

ARPS is the most common method of accessing the database. ARPS uses read-only access to the database.

The number of programs with write access to the database is limited. The nightly transaction file processing inserts, deletions and updates to the database. There is an interactive method to immediately update the database. It is only available to the data entry personnel.

The nightly transaction file processing paradigm of database updating keeps the database relative static. This is a benefit to some users, but does not allow timely access to new and/or updated data on a routine basis.

4.2.4. Data Selection Capability

A query is the method a user employs to pose a question and obtain an answer from a database. This section concerns only the selection criteria and the selected fields. The selection criteria is the means of specifying how to select data from the database. The selected fields are then available for further processing such as inclusion in a report or a data set.

4.2.4.1. Requirements

If data selection is done interactively, the user must be able to terminate the query prematurely. This means that the user should be able to terminate a query at any time. He should not be forced to wait for a fixed number of selected records to be displayed.

The data selection process should provide easy integration of the techniques used to search coded and narrative data.

The data selection process must allow the use of wildcards. This must not be limited to trailing wildcard specification; embedded or leading wildcards must be allowed. For example, if * is the wildcard character, *C, A*B, AB* and A*B*C must be legal.

The data selection capability must include the following constructs:

pattern generation:	concatenation
pattern recognition:	=, <> (not equal), BETWEEN, CONTAINING
value in a range:	>, >= (greater than or equal to),
	<, <=, AFTER (for dates), BEFORE (for dates)
relational operators:	AND, OR, NOT

The data selection capability must be able to select data where a field is coded as missing. For example, find all civilian ground accidents where the GRADE is missing (was not recorded).

4.2.4.2. Features and Problems in the Existing System

The data selection processing of the current ASMIS system is distributed among many programs. ARPS does the data selection for all ad hoc queries. All routine queries are handled with COBOL programs, each of which has its own data selection processing. Thus improvements to the data selection processing of ASMIS are very difficult because of the numerous encodings of the search strategy.

ARPS allows searching for a field coded missing by allowing searches of the form "GRADE = ." (1 or more blanks between the = and the .).

ARPS allows refinement of a previous query by use of the "hit list."

ARPS does not provide concatenation, the >=, <=, BEFORE, AFTER or NOT constructs. The lack of these does not affect functionality, rather it affects the ease of query specification. For instance, >= 5 can always be coded as > 4. BEFORE and AFTER are not currently necessary because ARPS permits manipulation of portions of dates, e.g., year or year and month. The addition of these constructs would provide more flexibility and would make the query specification easier from the user perspective.

4.2.5. Response Time

Response time is the time (wall clock) required to obtain the response to a query. This refers to the time used by the computer to respond to a computerized query and specifically excludes the time to translate a verbal query into its computerized form. It includes the time to generate the report specified by the query but not time to print it.

There are two kinds of queries: ad hoc and routine reporting. Routine reports are those queries which are run on a regular basis. Ad hoc queries are those which provide answers to one-time questions.

4.2.5.1. Requirements

Routine reporting does not require immediate response and thus, can be handled during times of low system utilization, i.e., overnight.

Response to an ad hoc query should be provided in an interactive mode. The system must be able to provide reasonable response times for each user under a typical work load of 10 users doing data entry and 20 users doing ad hoc queries. Reasonable response is in the 5 to 10 minute range (but 2 to 3 minutes would be better).

The time to respond to an ad hoc query is not constant. The system should be tuned to provide rapid response to often asked questions.

Over time, the type of questions that are frequently asked will change. The system must be flexible enough to be able to respond to these changes.

4.2.5.2. Features and Problems in the Existing System

All routine reports are run overnight as batch jobs.

Ad hoc queries are posed by both the external users and by the USASC staff. The ad hoc query capability is used extensively in the generation of Safety Center reports (e.g., the Annual Army Safety Report, the quarterly IPR).

The number of queries varies from day to day and week to week. The approach of a major report (Army Safety Report, IPR) deadline greatly increases the number of queries.

Prior to April 15, 1988, the response time for ad hoc queries was too long (30 minutes or more). The replacement of the IBM 4341 with an IBM 4381 has since reduced the response time dramatically. The response during May 1988 was in the 5 to 10 minute range. An increased workload, such as the approach of a major report, may change this.

4.2.6. Database Flexibility

Database flexibility is defined as having the facilities to alter the database structure or access methods to meet changing requirements with minimal impact on existing applications. A database is made up of fields and structures. A field is the smallest named unit of stored information, while a structure is a collection of related fields.

4.2.6.1. Requirements

The software system must provide a facility to manipulate the database structures easily and reliably. This includes adding, removing, and modifying both structures and fields within structures.

To allow the database to remain responsive as users' needs change, the software system must provide a facility to monitor and tune the database. Tuning the system includes changing indices and modifying system parameters. Changes can be made to improve database performance without changes to existing applications.

The software system must provide the ability to represent database structures in a clear and understandable manner, such as repeating groups and variable length strings. Personnel and property segments of the ground database would be represented as repeating groups, (multiple personnel and multiple pieces of property can be involved in a single accident), and narrative data would be represented as variable length strings (narrative data can consist of a variable number of characters).

The software system must provide data independence. The description of the database structure, its access methods and relationships among data must be independent of the application program(s) that uses it. Data independence enables the modification of the database structures to occur without modifying every application program which accesses the data. Only programs which use the modified field or structure would require changes. Changes to access methods and relationships among various types of data should be transparent to the existing applications.

4.2.6.2. Features and Problems in the Existing System

Currently, making file structure changes is difficult because it requires the modification of existing programs using the file. The structure of the data is known and used directly by the application programs. One change can set off a chain reaction of other modifications which have to be made.

Due to the difficulty in changing the structure of the data, the developers are occasionally unable to be responsive to the needs of the end users. Inconsistencies between programs can occur if all the programs that access the data are not modified. It can take more time and effort to modify applications then it took to write the programs initially.

4.2.7. Maintenance of the Database

The ASMIS databases contain a large number of fields (ground has approximately 200, aviation about 2000). The organization of this number of fields is not easy to comprehend. The existence of a high level document which shows the structures in the database, the links between various structures and includes the reasons for the decisions leading to the current structure would be valuable when modifications to the database are being proposed and implemented.

As the Army accident prevention system continues to function, changes to the data collected will be made. For instance, the codes for a particular field could be more finely delineated. The knowledge of exactly what was added and when this was accomplished is necessary to correctly interpret the historical data.

4.2.7.1. Requirements

A high level document describing the organization of the database should exist. As modifications are made to the database structure, this document should be updated.

A change control system should be used to track all changes made to the database. This will track changes in fields, structures, access methods, codes and textual translations for those codes.

4.2.7.2. Features and Problems in the Existing System

No centralized document for any of the four ASMIS databases exist.

In many instances, the knowledge of exactly when a change was made has been lost. The stated reason for not applying a complete suite of validations to the whole database is that the criteria for editing the field have changed.

4.3. USER INTERFACE

4.3.1. Data Accessibility

The Army accident database includes data for at least the past 17 years. The usefulness of the data decreases with its age and thus, the USASC has decided to store the oldest of the data off-line (on magnetic tape). This data is known as the historical data.

During this 17 years, the data collection forms have been changed to enable the collection of more detailed information. These changes have been reflected in the data sets in two different ways. The aviation data has two formats: pre-1972 and post-1972. For the ground database, the old format (pre-October 1980) has been translated to the new form. The old questions with no counterparts in the new version exist in a set of "old system" fields.

Data is accessible if the user can pose a query and receive a response.

4.3.1.1. Requirements

A user must be able to query any of the accident data that is available in the ASMIS database, regardless of its physical location. The historical data stored on magnetic tape should be accessible although that access may require more time than access to the current data which resides on disk.

Overnight response for a query requiring the off-line data will generally be acceptable. Occasionally, immediate access to off-line data will be required. Determination of whether a specific query should be run immediately will be an administrative task.

A user must be able to query any of the accident data regardless of the version of the form used to collect it. The database must be upwardly compatible and the start dates for new versions of the form must be clearly documented.

There should be no distinction between the accessibility of coded data and narrative data.

4.3.1.2. Features and Problems in the Existing System

The historical data is accessible by using the delayed or mailed option in ARPS. This allows these queries to be run overnight. At the end of second shift, the historical data is copied to a scratch disk and all jobs requiring this data are allowed to proceed. At the beginning of first shift on the following day, the historical data is removed from the system to allow scratch space for interactive users.

Queries using historical data can be handled during the day by reading the historical files directly from tape. For databases stored in one or two files, such as the ground accident database, this can be handled easily. The aviation accident database, which is stored in six data files, can not be handled as easily. A proposal for archiving aviation data exists and involves combining the aviation data stored in six files into one or possibly two files which could be stored on tape. This would require modifications to the ARPS program for aviation to allow it to read both the on-line and off-line formats.

The pre-1972 aviation data is not accessible through ARPS.

Queries involving searching the narrative data are restricted to the delayed or mail option of ARPS. This means that response to narrative searches are not available immediately, but require an overnight wait. Immediate response to a narrative search can be arranged by contacting the Safety Center.

Queries which only print the narrative but don't search the narrative do not require an overnight wait.

4.3.2. Query Specification

Query specification is the process of specifying the selection criteria, the selected fields and the subsequent processing to be done.

Options for subsequent processing include report generation or the generation of a subset of data.

4.3.2.1. Requirements

The user must have the ability to modify all portions of a query quickly, easily and independently.

The user must be able to input a list of case numbers as part of his query specification. He may have used the narrative portion of the record to determine cases of interest and thus there may be no way to specify the cases he desires by use of any other field or combination of fields.

The user must be able to save, modify and reuse queries.

The user must be able to abort the query specification process.

The user must be able to determine what will happen to a completed specification. He may choose to save and/or execute it. He should be able to specify the type of execution he wishes (immediate or delayed and the destination of the output generated).

4.3.2.2. Features and Problems in the Existing System

To facilitate query specification, a master library of queries (or portions of queries) is available (the PROCS for a particular database). A query from this library can be modified by specifying arguments before invoking the PROC. The displayed fields portion of the PROC can not be changed. The creation of a query for storage in this library is difficult and is accomplished by hand. This limits the number and complexity of queries in the library. Knowledge of this library is limited.

The method for saving a query is not known to all users. The saved query can be rerun, but not modified without extensive knowledge of the ARPS system and the computer operating system.

Queries can be modified during creation, but the method used to allow modification (use of EX, .EX, .AR) often requires that the user reenter correct information as well as the incorrect information.

4.3.3. Report Facility

The data selection process produces a data set consisting of one record for each case satisfying the selection criteria. Each of these records contains the selected fields specified in the query. Reporting is making this information available to the user in a easily readable form.

4.3.3.1. Requirement

The reporting facility will be able to produce the following kinds of reports.

- Case listing. A listing of all fields for each record selected.
- Columnar report. A columnar report is one in which the columns represent selected fields and the rows represent the records chosen from the database. No limits shall be placed on the sorting of the data for this report. Subtotaling and statistical summarization (e.g., sum, average, count, minimum and maximum) shall be available.
- Two dimensional matrix. A matrix shows values of one field versus the values of a second field.

The reporting facility will not limit the information to be reported to the width of the output medium (terminal screen or printed page). Rather multiple pages or screen will be used if necessary.

The reporting facility should print data in a user ready format. The user will be able to specify the format he desires. For instance, he can choose to display either the code used in the database or the textual translation of that code. He can control the pattern used in reporting dates, e.g., mmddyy, dd-mm-yy, etc.

The report generated shall include a statement of the query used to produce it.

4.3.3.2. Features and Problems in the Existing System

An ARPS report does not contain a statement of the query used to produce it. This facility is currently being added to ARPS.

ARPS reports are limited to the width of the terminal or the printed page. This results in multiple executions of essentially the same query. For instance, to get a matrix of two fields, each having more than six distinct values, the user must specify essentially the same query multiple times. To report different values of the horizontal field of the matrix, he must specify the value to be assigned to each of the six columns in the matrix.

4.3.4. Isolation of a Subset of Data

Some analyses take a number of days or weeks to complete. It is necessary to maintain a consistent subset of the data during the whole analysis.

Some analyses will require computational or display techniques not provided by the reporting facility.

4.3.4.1. Requirements

The user must be able to obtain a complete and internally consistent subset of the data.

The data subset shall include a statement of the query used to produce it.

4.3.4.2. Features and Problems in the Existing System

The various ASMIS databases have a field, SDATE which indicates the date the record was added to the database. Including a limitation on SDATE in the query protects the user from records added after the specified date. The user may also save a hit file (a list of case numbers) of cases selected and this file can subsequently be used to select the cases of interest. In either case, there is no protection against the updates applied to records which occur in the selected subset.

To produce large reports like the Annual Safety Report or the In Progress Report (IPR), an ASMIS database, e.g., ground, must remain constant. This is accomplished administratively by holding all updates until the report is completed.

The user can use another product to analyze the subset of interest. ARPS can write a report which can be used by another product (i.e., all PROCS named SAS . . . write data and no headings). The amount of data that can be transferred to the other product easily is very limited.

4.3.5. Alternative Computing Resources

Alternative computing resources are software products other than the ASMIS query processor and reporting facility. These facilities may reside on the USASC IBM mainframe or they may be available on another machine, a personal computer, for instance.

4.3.5.1. Requirements

Computing resources other than the ASMIS query processor and the reporting facility must be available to the users. These resources will provide:

- tools for more comprehensive analysis or display.
- the ability to manipulate an isolated data set.

The ASMIS system must provide the methodology to move data into these facilities. USASC must facilitate the loading of data into alternative software package(s). This includes moving the data, and establishing the field names and the textual translations associated with coded fields.

If the alternative computer resources exist on another machine, e.g., a personal computer, the USASC must also facilitate the transfer of data, field labels and code translations to the other machine.

4.3.5.2. Feature and Problems in the Existing System

ARPS can write a subset of data for use by another software product. SAS is used on the mainframe and various database products and spreadsheets are used on personal computers. ARPS can transfer only a limited number of fields of data (up to 132 characters of data) for each case meeting the selection criteria. The ability to transfer up to 512 characters is available to the DOIM programmers in batch mode. It is the responsibility of the user to establish the field names and the text for code translation.

Subsets of data are moved to personal computers for analysis. The technique for doing the download to a PC is documented but only in the PC software manuals and not in any Safety Center documentation.

4.3.6. Maintainable User Applications

A user application is maintainable if it can easily be adapted to accommodate change. Changes come from the computing system software, including the operating system and the database system, and from the application area, in this case, the Army safety prevention community.

There are several practices that can make application software maintainable. One is to isolate all software dependencies on the operating system and database system in a few routines. The applications should be written to incorporate good software engineering practices such as information hiding and modularity. Information hiding is the practice of keeping as much information as possible invisible from the users of the routine. For instance a good use of information hiding would be to have a single routine which reads the database and returns the next record. The user of the routine needs to know only about the record and possibly an end-of-file status flag. With only one routine having this knowledge, changing the method of reading the database is simpler (you change the one routine and relink all programs which use the routine).

Recording the changes is also important. Return to a previous version or reimplementation of a specific feature of a previous version is always a possibility. If a problem arises with the new production version of an application, knowledge of the changes made since the last production version can greatly reduce the time required to track down the error.

One program does not exist in isolation. It is important to understand the organization of the whole collection of application programs.

4.3.6.1. Requirements

Application software should be as independent as possible of the operating system and database system with which it communicates. If system dependent code must be used, such code should be isolated in a routine or a few routines and the details of dependency should remain inside those routines.

Common functionality should not be duplicated in multiple programs, rather that functionality should be in a routine by itself. This routine should then be used by all programs which need the functionality.

A method of tracking change to all code should be used.

A high level document describing the organization of the applications software should exist. As modifications are made to this organization of the applications software, this document should be updated.

4.3.6.2. Features and Problems in the Existing System

Application programs are not data independent. The description of the file structures and retrieval mechanisms is known and used directly in the application programs. If the structure changes, many programs must be modified. This work is expensive, time consuming, error prone and tedious.

Some of the fields within the data files are encoded and must be decoded to be of use to the user. Applications which access these fields each have a copy of the algorithms needed to access and decode these fields. If the decoding algorithms or the access method changes, then all the application programs which access these fields would need to be changed. If these access and decoding routines were made into procedures that could be placed into a common library to be used by the application programs, then only the procedures would need to be modified if the related data changed. The applications would only need to be relinked to make use of the modified procedures.

The code management facility, Panvalet, is available, but some of the code is not under its control.

No documentation of the whole application system exists. The information in chapter two of this document was gather by interviewing all the programmers and synthesizing the structure.

4.3.7. User Documentation

User documentation is a written description of the ASMIS database, presented from the perspective of someone wishing to use the database. Documentation is generally available in two forms, printed and on-line. A manual is printed documentation. On-Line documentation is a written description which is available on the computer system.

4.3.7.1. Requirements

On-line documentation shall be integrated into the various processes comprising the ASMIS system. This on-line documentation will include descriptions of both the process of query specification and the data itself.

On-line help must be organized so it can be successively disclosed to the user. At each level of disclosure, the user must be able to pick the portion of the help that he is interested in seeing next. He must be able to move up and down in the hierarchy of help.

User documentation shall also be available in printed form.

4.3.7.2. Features and Problems in the Existing System

On-line help is limited to a few questions within ARPS and concerns only the creation of a proper query. No on-line documentation is available for the database contents.

4.4. COMPUTING ENVIRONMENT

4.4.1. Data Security

Data security is the collection of protection mechanisms used to keep unauthorized users out of sensitive data while providing authorized users access to that data. Sensitive data is data which if it was released could cause harm to the individuals involved or to the Army. Example of sensitive data are name and social security number of individuals involved in an accident, names of aviation review board members investigating the accident.

4.4.1.1. Requirements

The USASC grants access to the databases within ASMIS on a need-to-know basis. Access to one database should not necessarily imply access to the others. For instance the safety community should not have access to the Drug and Alcohol Database and vice versa.

The USASC has established three levels of data access (general user, limited and Safety Center) as described in Section 3.7.3. No user should be able to circumvent his assigned access level and obtain data assigned to a more restrictive level.

No user should have access to files (particularly output from queries) belonging to another user.

The USASC must protect the database from intentional or unintentional destruction or corruption.

4.4.1.2. Features and Problems in the Existing System

Permission to access the Army accident data implies access to the aviation, ground and FECA databases. This is not necessarily a problem, but the possibility of restricting users to only one of these databases should be considered.

The data access levels described in Section 3.7.3 are only implemented within the ARPS program. A person wishing to avoid the limitation produced by the access level checking can circumvent the security ARPS provides by not using the ARPS program.

Once outside the ARPS program, access restriction is provided by passwords at the file level. Password protection at the file level is not being used consistently. Examination of a file list produced April 15, 1988 showed that of the thirty five files which comprise the four databases under the ASMIS system (the file names are listed in Section 3.2.1 through 3.2.5), only three files had file passwords. Thus, 90 percent of the files which make up the ASMIS databases are unprotected against deletion or corruption, whether intentional or unintentional.

When a new file is created, the creator can specify the file protection. If no protection is specified, the file has no file protection set. This lack of file protection allows a user to have access to files owned by other.

4.4.2. System Security

System security is the collection of mechanisms used to prevent access to computer system resources (e.g., ARPS, the COBOL compiler), by unauthorized personnel while providing access to the authorized users.

4.4.2.1. Requirements

The USASC uses administrative controls to evaluate the need-to-know of a potential user. A person with a need-to-know is granted access to the USASC computer by the assignment of a user identification and a password. The computer should validate the user identification and the associated password for each access to the computer system.

Periodic revalidation of need-to-know for all user should be required. Computer access for those without a current need should be terminated.

A user should have access to only the computing facilities necessary to complete his assigned task. He should not have access to facilities not necessary for his task. Typical ASMIS users need access to the database(s), the files they create and the analysis tools available on the system (e.g., SAS).

4.4.2.2. Features and Problems in the Existing System

User identification and password are validated by the computer system. A user can not change his own password, rather the change is handled administratively and the new password is provided either verbally or in written correspondence. This method lengthens the time required to close the security gap caused by disclosure of a password. It also provides an opportunity to compromise the newly assigned password. In fact, having a password in written form provides a continuing possibility for compromise.

Once a user has access to the USASC computer, he has access to the full power of TSO by simply choosing a menu item within ARPS. He now has access to the full range of facilities, including the COBOL compiler, the linkeditor, and ability to list, copy, and delete files without file password protection. Simply by knowing another person's user identification, he knows the high level qualifier on all that person's files and can delete, view or corrupt those files.

4.4.3. Data Backup

4.4.3.1. Requirements

USASC must provide backup of the data within the ASMIS system to insure the integrity of the data and to prevent loss. Frequency of backup should reflect the frequency of change of the data.

As long as data entry is handled by doing a nightly update of the database using the transactions created during the day, the backup of the data should precede the transaction processing.

The backup procedures must be easy to use and effectively incorporated into the routine operation of the Safety Center. The procedures must be efficient to minimize the impact on the production work.

USASC should maintain multiple versions of the backup for each type of data.

USASC should arrange and use storage in another building.

4.4.3.2. Features and Problems in the Existing System

Currently, backup of the data files comprising a database are the responsibility of the DOIM programmer(s) responsible for the database. Since portions of each database are handled by different programmers, the potential for loss of data integrity exists. The current division of responsibility does not present this problem, but other divisions could create problems.

4.4.4. System Backup

4.4.4.1. Requirements

USASC must provide backup of all system software. This includes all software from IBM and all third party software.

USASC must provide regular backup of all the files residing on the USASC computer to provide the ability to recover from disaster (e.g., disk crash, fire). The backup procedures must be easy to use and effectively incorporated into the routine operation of the Safety Center. The procedures must be efficient to minimize the impact on the production work.

USASC should maintain multiple versions of the system software backup(s) and of the full system backup.

USASC should arrange and use storage in another building.

4.4.4.2. Features and Problems in the Existing System

Currently, a weekly backup of the entire system is being done by doing several actuators a night on a rotating basis and by doing incremental backups on all other actuators nightly. Multiple versions are being maintained. These tapes are not currently being stored in another building.

4.4.5. Data Communications

Data communication is the ability for users external to the USASC to give input to and receive output from the USASC computer system. For most users, this access requires the availability of a communication line, a modem and a terminal (or a personal computer).

4.4.5.1. Requirements

USASC must provide world-wide access to the Army accident database. Dial-in lines as well as access through a military communications network should be provided.

Because of the world-wide access, USASC must minimize the time used for system backup and maintenance and maximize the time for user access. This allows users from other parts of the world access to the accident database at hours convenient to them.

To avoid limiting user access, the USASC must support communications to a wide variety of terminals. Consideration should be given to the whole range of terminals (from line-oriented devices such as a TI Silent 700 or a teletype, to IBM 3270 series terminals and personal computers emulating a variety of terminal types) when the types of terminals to be supported are chosen.

To avoid limiting user access, the USASC must support communications at a variety of baud rates. For the dial-in lines, 300-, 1200- and 2400-baud should be available.

4.4.5.2. Features and Problems in the Existing System

Currently, the USASC computer can be accessed through Defense Data Network (DDN) or by using commercial dial-in lines. The dial-in lines support 300- and 1200-baud communications.

The ASMIS query processor, ARPS, does all input and output in line-mode. Line-mode ("question and answer") can be supported on any type of terminal.

The quality of the communications over the dial-in lines is often poor (spurious characters often appear to both the user at the terminal and to the USASC computer). This results in garbled output for the user. If the spurious characters are going just to the user, then these characters are simply being added to his output, producing a cosmetic problem. If the spurious characters are going to the computer, the user is not able to specify his query correctly and thus his output will not be what he asked for. This causes many queries to be rerun.

Loss of carrier is another problem for the dial-in lines. When a carrier is lost, the user must reestablish his connection to USASC. Failing to make the reconnection in the limited time allowed (six minutes) will result in the query being re-entered and rerun from the beginning.

5.0 POSSIBLE IMPLEMENTATIONS

This chapter describes five possible implementations to provide access to the databases within the ASMIS system. The final section is a summary which relates the implementations to the functional requirements and identifies administrative issues such as cost of purchased software and level of training required.

In the proposed solutions described below you will find two recurring goals. One goal is to find a solution where the structure of the data is flexible. This will allow the database to be easily modified to reflect changes occurring in the Army safety community. A second goal is to provide a system with enough flexibility to be able to respond to the changes in the types of questions asked. This requires a method of changing the performance of the query processor.

Each section below will briefly describe a possible implementation. Obviously many features could be added to each solution. Included in each section is a description of the implementation, an identification of its unique features, a list of advantages, a list of disadvantages and a diagram showing an overview of the system. In the figures, icons have been used to represent the following types of structures. Figure 5.0 shows a sample of each icon.

- Organizational entities are groups of people. An example is the users.
- Computerized entities are structures which run on a computer. An example is the ARPS program.
- Stores of data. Typically this is a file on a computer. In these diagrams, use of the word database is not meant to imply anything about the form in which the data is stored. Rather database is used to identify the collection of Army Safety data.
- Paper output. An example is the report printed by ARPS using the mailed processing mode.
- Output formatted for paper but stored on the disk. An example is the report generated by ARPS using the delayed processing mode.

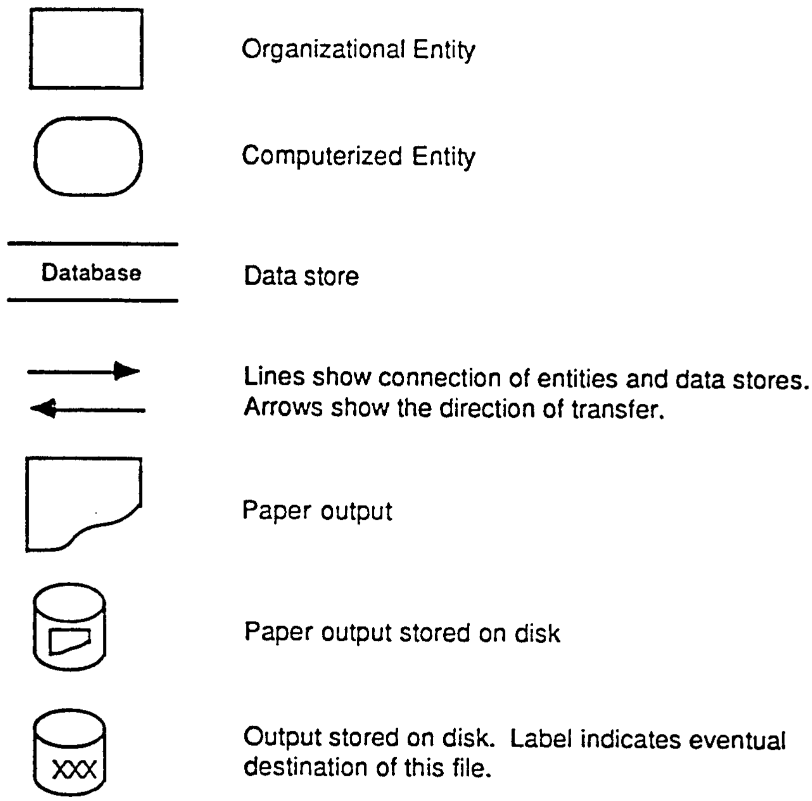


FIGURE 5.0 Legend for Other Figures

- Output formatted to be directly input to another computing resource. An example is information formatted for input into SAS, or a personal computer based product like Lotus 1-2-3 or DBASE III. The current ASMIS system provides only a primitive version of this type of output.
- Lines with arrows connect organizational entities, computerized entities, data stores and types of output. The arrow head indicates the direction of information transfer.

5.1. EXISTING ASMIS SYSTEM

This is the system currently being used by USASC. The file structure uses single keyed VSAM files. Figure 5.1 is a diagram of this system. For interactive queries, ARPS obtains the query specification from the user and using these criteria, gets the data from the database and displays the desired fields. Reports necessary for the routine reporting have been coded in COBOL.

5.1.1. Features

Given a query, all records which lie in the specified time period (e.g., fiscal year 1987) are sequentially searched to locate the records which meet the additional requirements provided in the user's query.

All queries for the same time period using the same type of data, for example, the basic and personal data from the ground database, take the same amount of computer resources. Ignoring the process of displaying the data to the user, the time to select 100 records from the data for a time period is the same as the time to select 5,000 or 10,000 records from the same time period.

For a given time period, the computer resources required to fulfill a query depends on how many of the types of data are required. For example, queries requiring just the basic data from the ground database take less computer resources than queries which involve the basic and personal data. This is because the portion of each ground database record which is not needed to evaluate the query is not expanded. For the aviation data, where the various types of data are stored in different files, this is also true because the data which is not required to evaluate the query is not read from those files.

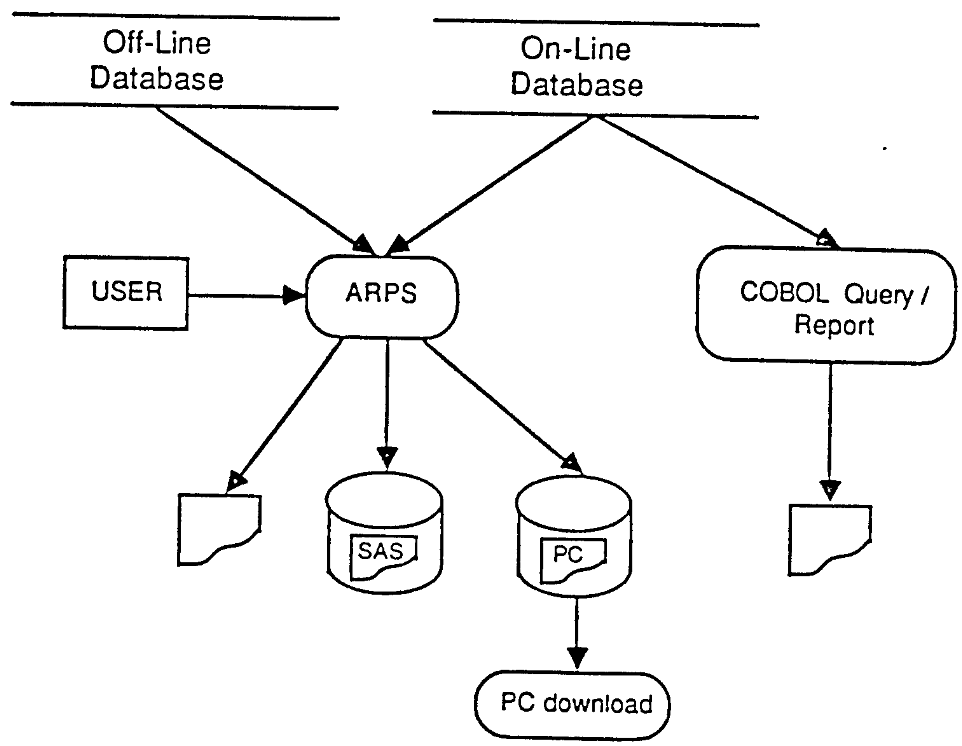


FIGURE 5.1 Existing ASMIS System

The historical data can be maintained off-line if desired. The penalty for maintaining this data on a sequential rather than a random-access medium, is searching all the data rather than the subset of data identified by the start and end date specified in the user's query.

5.1.2. Advantages

No software development is required to maintain the current position.

5.1.3. Disadvantage

This implementation lacks data independence. Changes to the structure of the data require change in all the programs (the ARPS program and the programs for the routine reports).

This implementation lacks data structure flexibility. A change to the data structure requires changes to the ARPS program and the programs for routine reports.

This implementation lacks flexibility in applications development. All applications are developed in COBOL. The Easytrieve package is no longer being used.

This implementation has limited reporting formats (caseprints, columnar listings less than 132 characters/line, and 2 and 3 dimensional matrices with limited horizontal and vertical size). It is difficult to transfer data from ARPS into another package such as SAS. It is difficult to transfer data from the IBM 4381 to a personal computer.

The only mechanism for integrating the various database (ground, aviation, FECA) is to duplicate data from one database in another database.

Control of access to protected data is distributed.

Little possibility for speed improvement exists, beyond what has already been done.

5.2. CENTRALIZED SINGLE USER QUERY PROCESSOR

This is a modification to the current ASMIS system. Basically two changes would be made in ASMIS. First, the query processor would be centralized.

All programs needing data from the database would obtain the necessary data from the query processor. Second, the types of output from the ASMIS system would be expanded to include output formatted to be directly input into another computing resource, for example SAS or a personal computer-resident spreadsheet or database management program. No change would be made to the file structure. Single indexed VSAM would still used. Figure 5.2 is a diagram of this option.

This implementation would be very similar to the current ASMIS system. The only difference would be that the process of obtaining a user's query criteria, obtaining the necessary data and generating an appropriate report to be returned to the user would be divided into two pieces. The generation of a report or a subset of the data to be input to another program would be isolated in another module. This isolation would allow all queries, both ad hoc and routine reporting, to use the same query processor. This isolation of the reporting module could allow the generation of multiple reports from a set of selected data.

5.2.1. Features

The method of accessing the data and the computer resources necessary to respond to a query would be identical to the existing ASMIS System (see Section 4.1).

The historical data could be maintained off-line if desired. The penalty for maintaining this data on a sequential rather than a random-access medium would be searching all the data rather than the subset of data identified by the start and end date specified in the user's query.

The types of output would be expanded to allow output specifically formatted for use by other computer packages.

5.2.2. Advantages

It would be possible, but not easy, to change the file structure in the database. To change the file structure would require change to the query processor and to the data entry process (which is not shown in Figure 5.2).

The query processor would be centralized. Any speed improvements which could be obtained would apply to both the ad hoc and routine reporting queries.

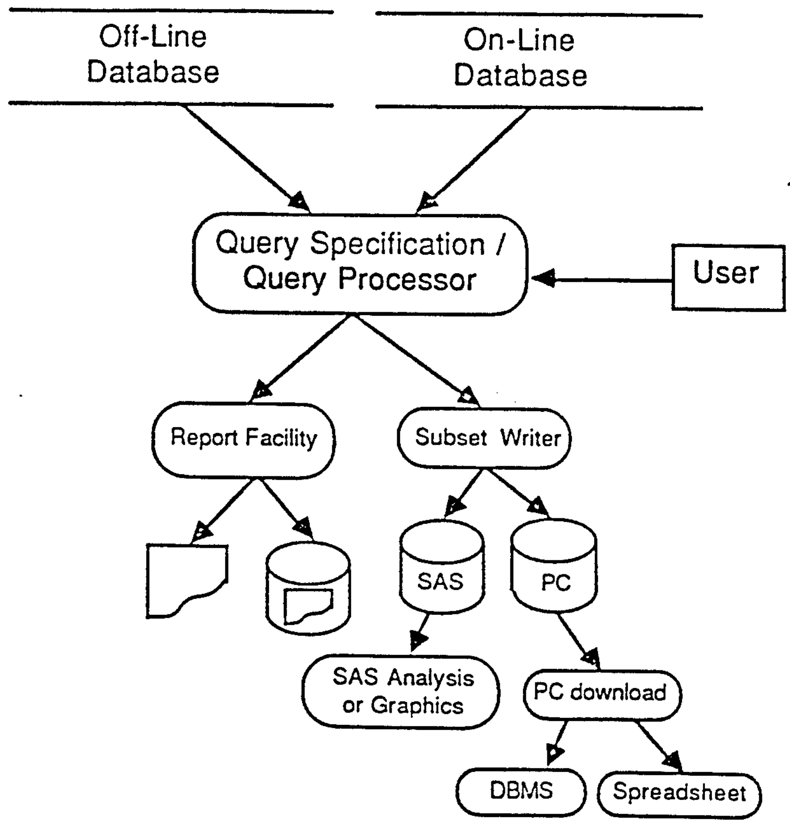


FIGURE 5.2 Centralized Query Processor

It would be much easier to transfer data from this implementation into another package such as SAS or to a personal computer. Purchase of additional software is not required for this implementation.

5.2.3. Disadvantages

This implementation would still lack data independence. Changes to the structure of the data would require change to the query program and to the data entry process.

This implementation would lack flexibility in applications development. All applications would continue to be developed in COBOL.

This implementation would have limited reporting formats because of the difficulty in developing a more extensive list of options.

The only mechanism for integrating the various databases (ground, aviation, FECA) would be to duplicate data from one database in another database.

Control of access to protected data would still be distributed.

Centralization of the query processor provides the opportunity to speed up the query processing. Unfortunately, the VSAM implementation used in this system lends itself to the use of only single indexes; multiple indexes can not be used together in responding to a query. Little speed improvement could be gained.

This implementation would require training for the programming staff and for the users. The programming staff would need to learn to integrate the centralized query processor into their programs. The paradigm for generating a report would be to use the query processor to extract the required data followed by a program to compute and display the data appropriately. The users would need training to take advantage of the increased facility to transfer data from the query processor/reporter into a statistical analysis package or down to their personal computer.

5.3. CENTRALIZED MULTIPLE USER QUERY PROCESSOR

This version would be a very different implementation of ASMIS. The difference is that the process of obtaining a user's query criteria, obtaining the necessary data and generating an appropriate report to be returned to the user would be divided into three pieces. The first step, obtaining the user's query criteria (called query specification in Figure 5.3) would resemble the current ARPS program. There would be a conversation between the program and the user to determine the criteria for selection of the data. This query criteria would then be passed to another process, the query processor, which would read the database for appropriate records. As in the Centralized Single User Query Processor, the generation of a report or a subset of data for input into another program would be isolated in a separate module.

This query processor would be structured differently than the corresponding piece of the current ARPS program. This processor would be capable of evaluating multiple queries for each pass over the database. A real life example of how multiple requests can be handled more efficiently by a single entity will clarify the concepts used. Consider the situation where all people in an office go to lunch using the following rules:

- No two people can go to lunch at the same time.
- Each person goes to town to obtain his lunch and returns to his office to eat it.
- There is only one car for the use of the whole office staff.

One implementation of the lunch hour is to have each person take his turn going to lunch. Each person drives to town, orders his lunch and returns to work with lunch in hand. If the time to obtain one person's lunch is 15 minutes, the time for 10 people to obtain lunch is 150 minutes.

A second implementation of the lunch hour is to have one person collect orders from all the others, drive to town, order all the lunches and return with them. The time for 10 people to obtain lunch is now much less than 150 minutes, but somewhat more than the 15 minutes necessary to obtain a single lunch. One further restriction is also obvious: to gain maximum speed in

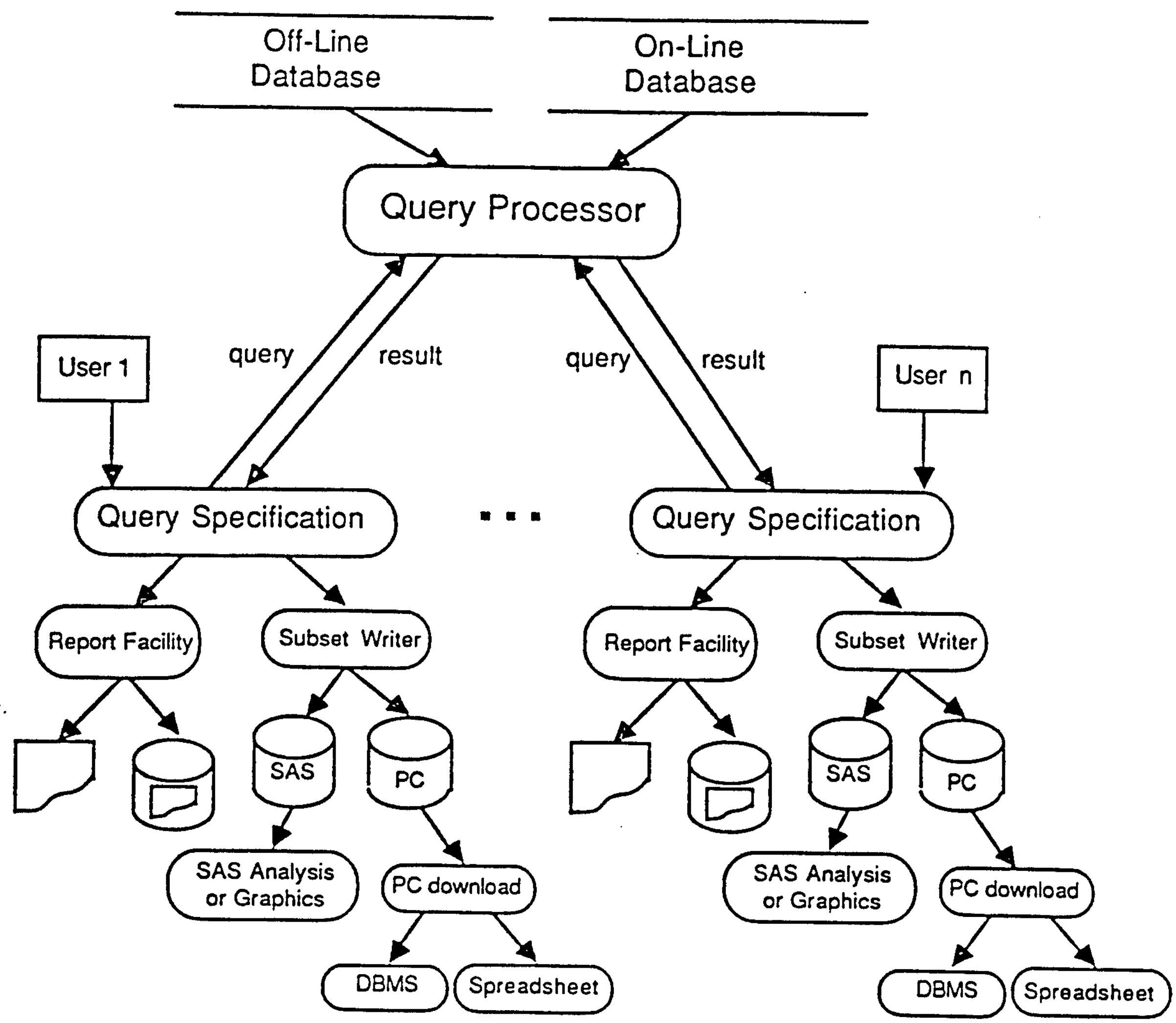


FIGURE 5.3 Multiple User Query Processor

getting lunch, all lunches come from the same restaurant. The difference between the lunch example and the Centralized Multiple User Query Processoris that lunch occurs once a day while the query processor continually repeats its job, the searching of the ASMIS database.

5.3.1. Features

More than one query would be fulfilled for each pass over the database. The query processor would be a task which repeatedly made passes over the whole database. At the start of each pass, all queries which were completely specified would be passed to the query processor. When the current pass over the database was complete, the records to satisfy all these queries would have been selected. The process of reporting these to the user (as a report or a subset of data for input into another program) would now be left to do.

All queries for a database would take the same amount of time. There would be no advantage to getting data from only a subset of the database. For instance, retrieving data from just the basic portion of the ground database would take as much time as retrieving data from the basic, personal and equipment portions. Retrieving data from a short period of time, for example one year, would take as long as retrieving data from all years.

There would be no penalty for maintaining data on a sequential medium because all records in the file are read for each pass. Thus, off-line data is easy to integrate.

Because of the restriction that any request takes as long as the longest request, this solution would only be suited for the case where more work needs to be accomplished than there is available computer time. It would perform better than the Centralized Single User Query Processor or the current ASMIS System on a saturated system. It would perform poorly on a nonsaturated system.

The time to fulfill a request would still appear to be variable to the user. The user would see the sum of two times, the time his query must wait for the start of a pass over the database and the time to make that pass.

5.3.2. Advantages

This version provides a different query processor which would provide a speed improvement by handling many queries at once. A estimate is that ten queries could be handled in the time to do two queries using the Centralized Single User Query Processor or the current ASMIS system. Because the machine was fully saturated, this modification would have improved the ASMIS system on the IBM 4341 remarkably.

Like the Centralize Single User Query Processor it: 1) would be possible, but not easy to change the file structure in the database, 2) would be easy to transfer data from this implementation into another package such as SAS or to a personal computer, and 3) would require no additional purchased software.

5.3.3. Disadvantages

This version would still have all the disadvantages of the Centralized Single User Query Processor. It would lack data independence, would lack flexibility in application development, would have limited possibility for file structure changes, would have limited reporting formats, would have no good mechanism for integrating data from different databases, and the control of access would still be distributed.

The response time for the user would be longer. In addition to waiting for his query to be evaluated and the data to be returned, he would have to wait for the start of the next pass over the database.

The software for the multiple user query processor is more complicated than the software for the centralized single user query processor or for the existing ASMIS system.

Like the Centralized Single User Query Processor, this implementation would require training for the programming staff and for the users. The programming staff would need to learn to structure their programs to integrate the centralized query processor. The users would need training to take advantage of the increased facility to transfer data from the query processor/reporter into a statistical analysis package or down to their personal computer.

5.4. DATABASE MANAGEMENT SYSTEM (DBMS)

This is an entirely new implementation. The current ASMIS system would be replaced with a system designed on top of a commercially available database management package.

One important feature of a DBMS is its ability to provide answers to simple questions quickly while still being able to answer very complex questions. Comparing the DBMS to the lunch example in the previous section provides some insight into this feature of a DBMS. Using a DBMS is like using the first implementation of the lunch hour (each person goes to get his lunch separately), but the source of the lunches is now varied. Some people just go down the hall and retrieve their lunch from the refrigerator and are thus very fast. Others drive just a few blocks and return fairly quickly. A few people drive clear to the other end of town, taking substantial time to obtain their lunch.

Another important feature of a DBMS is the flexibility to be able to respond to changes in the types of questions being asked. An office filing system will provide an example of a real world problem and the solution represented by the DBMS. In a very small office everyone knows the filing system and has access to the files. As the office increases in size, the open access to the files presents a problem. The filing scheme can not be changed because too many people need access and communicating the changes to the filing system would be too difficult. The eventual solution is to isolate the filing system and hire a file clerk to do all access to those files. Now the change to the filing system is easy. Only one person, the file clerk, needs to know the changes. In fact, no one but the file clerk would be aware that the underlying structure of the filing system had even been changed. The data in the DBMS is like the filing system and the query processor is the file clerk.

A DBMS provides a non-procedural method for access to data. Non-procedural means that only the desired result must be specified, not the detailed steps necessary to obtain that data. The typical communication between a boss and his secretary is often non-procedural. The boss may simply ask that something be done, but does not specify how it will be done.

5.4.1. Features

The time to respond to a query is dependent on the query. Simple queries will receive a response in a short time; more complex queries will receive response in a longer time. The time to respond to a query is dependent on three things, whether the query can take advantage of fields which have been indexed, the number of records retrieved from the database, the complexity of the query. A query is complex if it requires data from more than one portion of the database. For instance, a query which needs the personal and basic data from the ground database is more complex than a query which requires only basic data. Queries which use fields in the selection criteria which have been indexed in the database will receive more rapid response than queries which do not use indexed fields in the selection criteria.

5.4.2. Advantages

This implementation provides data independence. Data independence means that the description of the database structure, its access methods and relationships among data are independent of the application program(s) that use it. Modifications to a data field would not require changes to applications which did not use that data field. Even applications which use the data field would not necessarily need modification. They would, of course, require review. Adding new fields of data would not affect applications except those that should be expanded to include the new data.

This implementation provides data structure flexibility. Fields can be added, modified or removed from the database with minimal impact on existing applications.

This implementation provides application development flexibility. Commercial software is available to assist in the development of applications. Thus, applications could be developed in COBOL or in a fourth generation language. A fourth generation language is non-procedural: you specify what you want, not how to get it. COBOL allows access to the same type of non-procedural query capabilities but provides the ability to perform operations on the retrieved data that are beyond the capabilities of the fourth generation language. This would facilitate the generation of additional reporting formats.

This implementation has enough flexibility to be able to change the performance of the query processor to match the changes in the types of questions being asked. Two types of change are possible. First, individual fields can be indexed to speed access. Second, the data can easily be reorganized to reflect the current usage.

The DBMS provides access to all data and thus can provide a single location to control access to protected data. With the previous implementations, you could avoid the query processor and read the data file(s), thus escaping the protection mechanism of the query processor. With a DBMS, access control is built into the DBMS. No processes can avoid its access control.

The DBMS implementation provides an easy way to integrate data from the multiple ASMIS databases. No duplication of data is necessary.

This implementation facilitates transfer of data to other packages or to a personal computer.

5.4.3. Disadvantages

Performance can be a problem. This problem is one of the subjects of the next chapter of this document.

To convert the ASMIS system to a DBMS-based implementation would involve a substantial development effort.

To effectively use the DBMS would require considerable training. Each programmer would need training concerning the concept of a DBMS and the new tools available. A database administrator (DBA) and a DBA backup should be named. The DBA would be responsible for overseeing the development, maintenance and tuning of the database. The DBA and backup would need considerable training to gain the skills necessary to structure and tune the database. The users would also need training.

This implementation can require the purchase of software. The initial cost and yearly maintenance costs can be considerable. This topic is considered in the next chapter. (Note, however, that the recommended DBMS is available to Army installations at no cost).

Data stored in a DBMS occupies two to three times as much disk space as data stored in a sequential file.

Data in a DBMS must be stored on a random-access medium (see Figure 5.4). Data can be archived to a sequential medium, such as magnetic tape, by extracting the data from the database. This data can then be deleted from the database, thus reducing the on-line space requirements. Providing convenient access to this old data requires reloading the data into the DBMS again before it can be accessed. This technique will not work for ASMIS because of the relatively short turn around time for questions relating to very old data.

No DBMS supports the use of data on sequential medium. To have data on sequential medium means that an alternative access method must be provided. For ASMIS this would mean a DBMS for the on-line data and an entirely different system for the off-line data. This strategy is not recommended.

5.5. DBMS WITH MULTIPLE USERS QUERY PROCESSOR

This implementation would provide a Multiple User Query Processor like the Centralized Multiple User Query Processor, but it would be based on a database management system. Thus it would provide most of the flexibility of the database management system while gaining the speed of the multiple user query processor.

As with the Centralized Multiple User Query Processor, this implementation provides the greatest benefit on a saturated system and the least benefit on a non-saturated system.

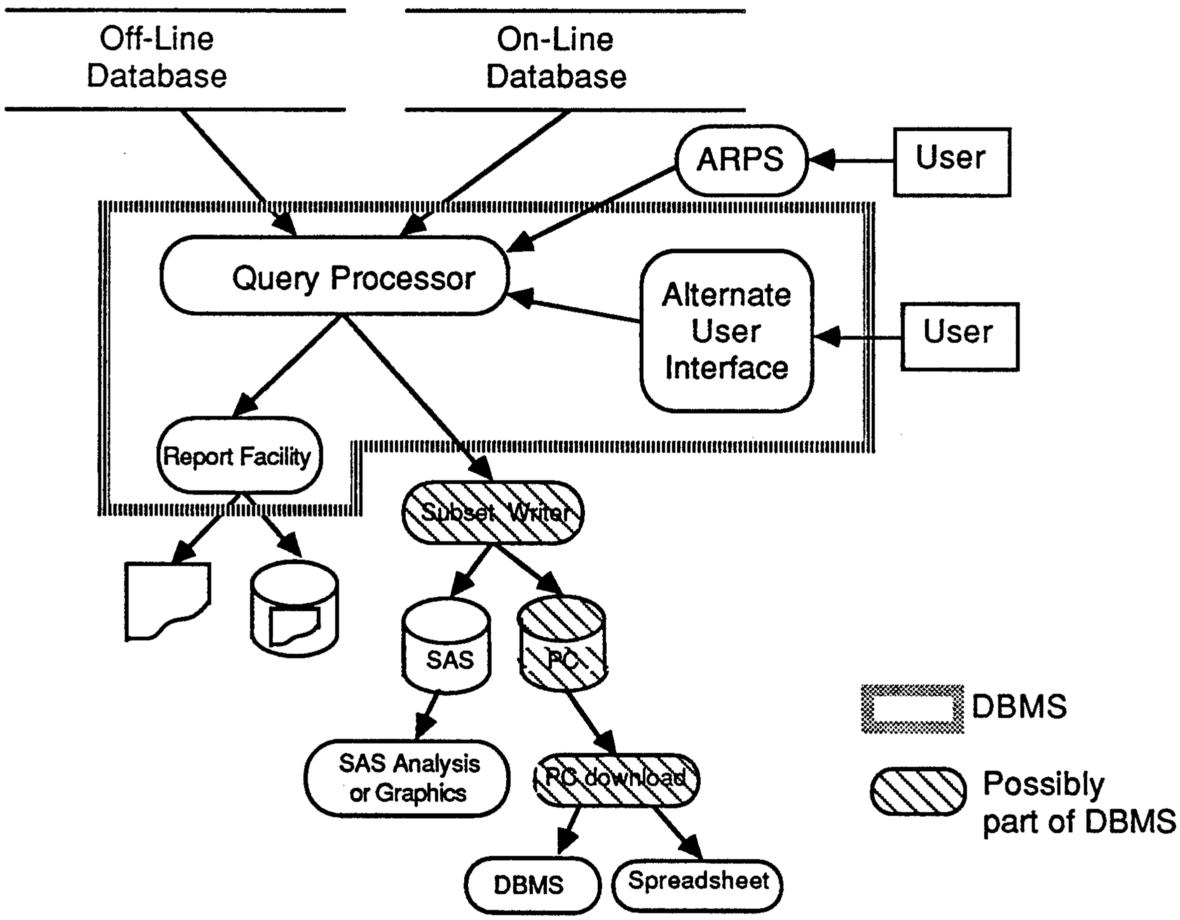


FIGURE 5.4 Database Management System (DBMS)

5.5.1. Features

Just like the Centralized Multiple User Query Processor, more than one query could be fulfilled for each pass over the database and all queries would take the same amount of time.

Unlike the Centralized Multiple User Query Processor, this implementation would require multiple copies of the data (see Figure 5.5). The data would be maintained in the database management system and the DBMS would be the means of providing data structure flexibility. A copy (outside the DBMS) of this structure would then created and used by the Query Processor. This would add to the storage requirement for the database.

The copy of the data which is maintained outside the DBMS would be easily generated from the DBMS. Thus changes to the data structure would be relatively easily reflected in the data used by the query processor. This would be better than the Centralized Multiple User Query Processor where change to the data structure are difficult, but less optimal than the DBMS solution where the change needs to be made only once.

The query processor itself would be specially constructed so that all knowledge of the structure of the data file is housed outside the query processor in some setup file. Thus no reprogramming of the query processor would be necessary to change the structure of the data (recompilation and link-editing could be required).

There would be no penalty for maintaining data on a sequential medium because all records in the file are read for each pass. Thus off-line data would be easy to integrate.

5.5.2. Advantages

This implementation would provide data independence. Changes to the data structure would be made in the copy inside the DBMS. Depending on the change, the copy outside the DBMS could need to be regenerated. If fields were added to or deleted from the database, the copy would need to be regenerated. If changes were made to an existing field (for instance, addition of new codes), no regeneration would be necessary. In any case, the process

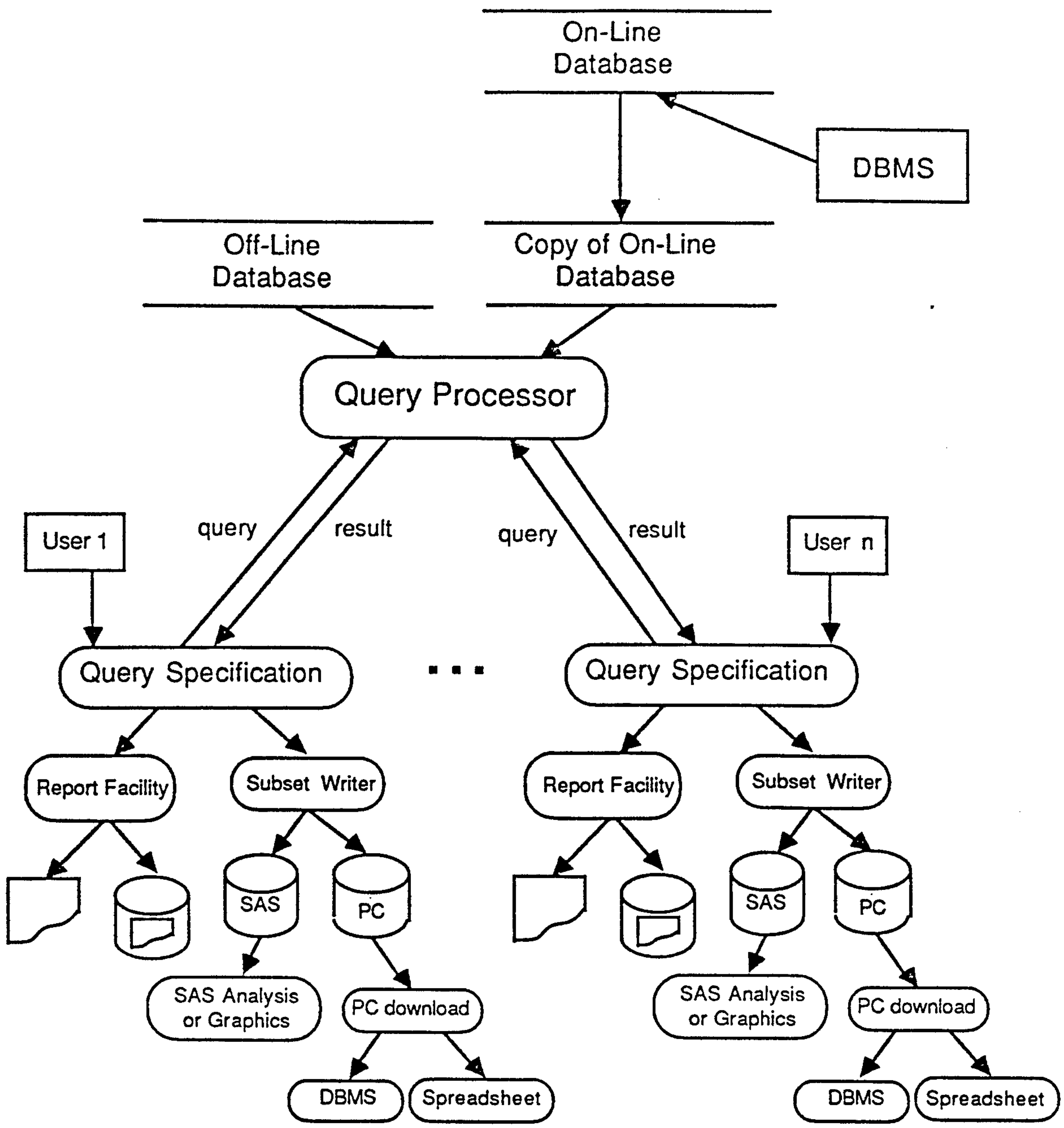


FIGURE 5.5 **DBMS with Multiple User Query Processor**

of regeneration would be automated and thus require computer resources (an overnight run), but little personnel time.

This implementation would provide data structure flexibility. Fields could be added, modified or deleted from the database with minimal impact. For some types of changes, the copy of the data outside of the DBMS would need to be regenerated.

This implementation would provide a different style of query processor which would provide a speed improvement by handling many queries at once. It would provide a means of responding to more queries on a saturated system.

For the users, this implementation would mean less training than the DBMS-based implementation and about the same training as the two centralized query processor implementations (described in Sections 5.2 and 5.3).

5.5.3. Disadvantages

This implementation would perform poorly on a non-saturated system.

This implementation would require more disk storage than the DBMS implementation.

This implementation would limit the application's flexibility. Certainly the options for developing applications code would not be as great as those under the DBMS. During the history of ASMIS, the problem has been the interactive queries not the routine (batch) processing. This suggests that the routine reports could still be produced using the data managed by the DBMS. In this case, the full set of tools associated with the DBMS would be available for use in routine reporting and thus the effect of this disadvantage would be reduced.

Like the DBMS implementation, this implementation would require the procurement of the database management software.

This implementation is more difficult than the DBMS implementation. It would require more effort to convert the current ASMIS system and require more training for the DOIM staff.

5.6. SUMMARY OF POSSIBLE IMPLEMENTATIONS

The previous five sections proposed implementations for both saturated and non-saturated systems. Table 5.1 identifies which implementations are appropriate for saturated and non-saturated environments.

TABLE 5.1. Appropriateness of Possible Implementations on Saturated and Non-Saturated Systems

	ASMIS	Centralized Single User	Centralized Multiple User	DBMS	DBMS with Multiple User Query
Performance on Nonsaturated System	Ok	Ok	Poor	Ok	Poor
Performance on Saturated System	Poor	Poor	Ok	Poor	Ok

The five possible implementations were evaluated against the functional requirements identified in Chapter 4. The following functional requirements were judged to have no effect on the choice of implementation.

- Data Acquisition (see Section 4.1.1). Nothing in the implementations will affect how the data is collected.
- Data Entry (see Section 4.1.2). The main issues in data entry, checking the form for completeness for the specific accident type and maintaining consistency between old and new data, are not effected by any of the implementations.
- Data Selection (Section 4.2.4). None of the implementations change the desirability of: 1) a full set of data selection constructs and 2) easy integration of coded and narrative data.
- Maintenance of the database (see Section 4.2.7). Nothing in any implementation will affect the desirability of a high level document describing the database and a complete history of what changes were made and when.

- Query Specification (Section 4.3.2). None of the implementations change the desirability of: 1) improved facilities to save, modify and reuse queries, 2) specification of case numbers as part of the selection criteria (use of the "hit list").
- Maintenance of user applications (see Section 4.3.6). Maintenance of a history of changes to the application programs is desirable in any implementation chosen. A high-level description of the organization of the applications software is desirable independent of which implementation is chosen.
- User documentation (see Section 4.3.7). Adequate documentation, presented from the perspective of someone wishing to use the database, is required for any implementation.
- System security (see Section 4.4.2). None of the implementations change the need for improved system security.
- Data backup and system backup (see Section 4.4.3 and 4.4.4). None of the implementations change the necessity of adequate backup for both the database contents and the its environment.
- Data Communications (see Section 4.4.5). None of the implementations has any influence on the communications for off-site users. Improved communications would benefit any implementation.

Table 5.2 compares the possible implementations using the remaining functional requirements. To provide a means of summarizing the relative goodness of an implementation, each possible implementation was evaluated for each functional requirement was also rated using a 1 to 5 scale. A "1" means that the implementation has a low rating for this functional requirement. Examples of a "1" are "uses most disk space," "provides fewest report options," and "maintenance is hardest." A "5" was given for implementations which performed best, e.g, "uses least disk space," "maintenance is easiest," etc. Ratings of 2, 3 and 4 were used to represent increments between the best and the worst. The number below the functional requirement indicates which section contains a description of the functional requirement.

Table 5.3 rates the five implementations for a number of administrative issues which were not covered in the functional requirements. Table 5.4 contains a final total which represents the sum of the functional requirements rating and the administrative rating.

TABLE 5.2. Comparison of Each of the Five Implementations Versus the Functional Requirements Where Differences Exist

	ASMIS	Centralized Single User	Centralized Multiple User	DBMS	DBMS with Multiple User Query
Data Independence (4.2.6)	1 Lacks	1 Lacks	1 Lacks	5 Has	5 Has
Data Structure Flexibility (4.2.6)	1 Lacks	2 Limited	2 Limited	5 Has	5 Has
Applications Development Flexibility (4.3.6)	1 Lacks	1 Lacks	1 Lacks	5 Has	4 Maybe Limited
User Applications Maintenance (4.2.7)	1 Harder	2 Hard	2 Hard	5 Easiest	4 Easy
Query Processor (4.3.6)	1 Distributed	5 Centralized	5 Centralized	5 Centralized	5 Centralized
Flexibility in Reporting (4.3.3)	1 Lacks	1 Lacks	1 Lacks	5 Has	4 Maybe Limited
Transfer of Data to Other Packages (4.3.4-4.3.5)	1 Hard	5 Easy	5 Easy	5 Easy	5 Easy
Data Integration Aviation, Ground, FECA (4.2.2)	1 No	1 No	1 No	5 Yes	1 No
Handle Off-line Data (4.2.2)	5 Yes	5 Yes	5 Yes	1 No	5 Yes
Ability to Gain Speed (4.2.5)	1 Little Adjust	1 Little Adjust	5 High Leverage	3 Some Adjust	5 High Leverage
Data Security/ Access Control (4.4.1-4.2.3)	1 Fragmented	1 Fragmented	1 Fragmented	5 Single Location	1 Fragmented
Total for Functional Requirements	15	25	29	49	44

TABLE 5.3. Comparison of the Five Implementations Versus Issues Outside the Scope of the Functional Requirements

	ASMIS	Centralized Single User	Centralized Multiple User	DBMS	DBMS with Multiple User Query
Total for Functional Requirements	15	25	29	49	44
Total Administrative	35	27	24	12	10
Final Total	50	52	53	61	54

TABLE 5.4 Summary of the Comparison of the Five Implementations Versus the Functional Requirements and Administrative Issues

	ASMIS	Centralized Single User	Centralized Multiple User	DBMS	DBMS with Multiple User Query
Software Cost	5 Low	5 Low	5 Low	1 High	1 High
Development Required	5 No	1 Yes	1 Yes	1 Yes	1 Yes
Development Cost	5 Low	4 Moderate	3 Moderate+	2 High	1 High+
Level of DOIM Training	5 Low	4 Moderate	3 Moderate+	2 High	1 Higher
Level of User Training	5 Lowest	4 Low	4 Low	1 High	4 Low
Type of Staff Required	5 No Change	4 Little Change	3 Some Change	2 High Change	1 Higher Change
Disk Space	5 No Change	5 No Change	5 No Change	3 More Needed	1 DBMS Plus
Total Administrative	35	27	24	12	10

6.0 EVALUATIONS

This chapter contains descriptions and results from the evaluations that were done in the process of preparing a recommendation for the future implementation of the ASMIS system. The computer system was evaluated to document its current status and project the effect of the installation of a commercial database management system. The style of queries currently being run through the ARPS program was analyzed and modifications to the criteria for the selection of a database were made based on this information. The various database management products were reviewed to select the one or ones which will provide the proper tools.

6.1. PERFORMANCE ANALYSIS OF THE CURRENT SYSTEM

6.1.1. Summary and Conclusions

This study was designed to obtain a detailed understanding of system behavior during first-shift operations. The results are based on performance data collected during a 75-minute interval each morning and afternoon for a week (5/23-27/88, 0830-0945 and 1300-1415). During these periods, the load consisted of an average of 17 TSO users (max 24), plus at least one batch job about half the time.

The performance data allowed us to discriminate several distinct classes of activity (e.g., batch jobs, short/medium/long TSO transactions). It does not allow us to tell for sure what tasks were being performed in each class. However, analysis of the query patterns indicates that most of the load (50 to 80 percent) is caused by ARPS queries. The remainder is due to program development, overhead and data entry. Data entry by itself comprises less than 3% of the load.

The system has substantial unused capacity at present. Utilizations for the four most heavily used resources are: cpu, 34 percent; disk channels, 27 percent; ASCCAT and SYSMT1 disks, 24 percent each. These utilizations indicate that the system could comfortably support double the current load, with no software changes. However, tripling the load would completely saturate the system and would not be feasible.

The hardware configuration appears well-balanced at present. It is clear that usage has increased since the 4381-13 upgrade. Under the current load, the older 4341-11 would have been 117 percent saturated.

Utilizations measured with the current ARPS software cannot be used directly to predict the effect of changing to an indexed database system. However, they do provide absolute measures for some system resources. In particular, the measurement data indicate that the system can sustain about 40 I/O requests per second for a single disk, and about 80 I/O requests per second across all disks combined. These numbers can be used to evaluate performance limits on I/O intensive database applications.

Three cautions must be observed in using these conclusions:

1) System performance outside the measurement periods may be substantially different from the above results. In particular, we made no attempt to characterize the nighttime batch load.
2) The entire measurement week may have been atypical. We know that the disk system was giving problems that week, and members of the USASC staff indicated that the load seemed somewhat lighter than normal.
3) The load for short periods may be substantially higher than the averages reported above. For example, the cpu utilization averaged over individual 15 minute periods was observed to be as high as 84 percent. Without software changes, substantial increases in throughput would require shifting work from busy periods to idle ones.

6.1.2. Procedure

To accurately estimate system behavior, we needed data that covered representative periods and was fairly detailed. We chose a 15-minute measurement interval to give an adequate level of detail, then picked the measurement periods, 0830-0945 and 1300-1415, to cover "typical" workload levels while limiting the amount of data to what we could afford to handle.

Performance data was collected using IBM's standard Resource Measurement Facility (RMF). This data includes the utilization of system resources (cpu, disks, channels), number of transactions, response time statistics, swap counts, etc. RMF divides some of the data into specific classes of workload. For example, transaction counts, cpu usage, total I/O usage, and response time statistics are reported separately for batch jobs and for TSO transactions in each of four different classes, determined by transaction length. The remainder of the data, such as disk utilizations, are aggregated across all classes.

There were several steps in analyzing this data:

1) Recognizing and removing measurement intervals that clearly did not reflect normal operations (e.g., disks hung).

2) Deriving meaningful resource utilization measures for situations that RMF doesn't handle well (e.g., logical channels having multiple physical channels).

3) Developing meaningful aggregations across classes. (RMF uses 22 classes, of which only 6 turned out to be meaningfully distinct.)

4) Constructing a "queuing network model" (QNM) to represent interaction between the workload and the computer system.

5) Evaluating the QNM to obtain predicted response time, and validating against the RMF data.

For the current analysis, most of the important results come out of Steps 1 through 3. The QNM developed in Steps 4 and 5 serves as a valuable crosscheck that no significant effects were missed. In addition, the QNM can be used to predict the effect of changing the hardware configuration and/or workload level. The current analysis does not require such detailed predictions.

Most of the analysis was done using standard tools from the Unix environment, especially the 'awk' language and 'sed' utility. Evaluation of the final QNM was done using MAP, a performance analysis package from Quantitative Systems Performance, Seattle, Washington.

6.1.3. Detailed Findings and Discussion

This section discusses the data and interpretations in some detail. It is intended as backup material for the major conclusions presented above.

Initial examination of the RMF data showed two occasions that clearly did not reflect normal operations. One occasion, covering the entire afternoon slot of 5/25, showed very strange disk behavior. (The disk service times were 10 times larger than normal, and the distribution of channel activity was uniquely skewed.) Jim Hayes of the USASC suggested that this might be one of the periods when the disk system was having hardware problems. The other occasion was a single 15-minute interval in the morning of 5/27, when the statistics for intermediate-length TSO transactions were 1000 times longer than those for all other intervals. We suspect this was due to some system error causing a single wild datum in the underlying RMF/SMF data. The data corresponding to both of these occasions was simply omitted from further analysis. This left valid data covering a total of 10-1/2 hours.

6.1.3.1. Workload Characterization

RMF divided the workload into 22 classes, most of which are various overhead functions. After examining the data, we decided to represent the workload as six aggregated classes:

TSOS - Short TSO transactions (typically 0.1 second)
TSOM - Moderate length TSO transactions (typically 1-10 seconds)
TSOL - Long TSO transactions (typically 100 seconds)
BATCH - Batch jobs
CICS - CICS subsystem
MISC - Miscellaneous system overhead

The overall statistics for these classes were:

Name	Cpu Share	I/O Share	Transactions	Average Response Time	Swaps per Transaction
TSOS	8%	5%	19289	0.13 sec	1.0
TSOM	41%	57%	5430	9.3 sec	1.3
TSOL	10%	5%	17	411 sec	37.4
BATCH	28%	21%	155	122 sec	.1
CICS	<3%	<1%	?	?	?
MISC	11%	11%	-	-	-

The RMF data does not allow us to tell for sure what tasks were being done within each of the classes. However, we can make some educated guesses.

First, note that a TSO transaction represents the period that starts with completion of user input (i.e., carriage return) and ends when the program has logically produced its final output and is ready for more input. For transactions producing only a little output, the "response time" is proportional to the machine resources needed to process the transaction. If a transaction produces substantial output, especially on a slow communications line, then it will be delayed while intermediate output is transmitted to the user. In this case, the "response time" will be disproportionately long, compared to the machine resources required, and will be accompanied by a large swap count.

Class TSOS, short TSO transactions, is attributable to ARPS setup commands, as well as miscellaneous TSO commands. This class includes very little if any retrieval activity.

Class TSOM, moderate length TSO transactions, corresponds well with ARPS queries in which substantial amounts of data are being retrieved, one block at a time (e.g., "YOU HAVE DISPLAYED 100 LINES. . ."). Each block corresponds to one transaction. In this case, only a few seconds of machine time is required per transaction, and output delays (swaps) are limited because limited output is being produced per transaction.

Class TSOL (long TSO transactions) probably contains two types of activity. The first type is typical of ARPS retrievals that find very little data. In this case, minutes of machine time are consumed, the response time is proportional to machine usage, and there is very little swapping. The other type consists of sending voluminous output to the user, without pausing after each block. In this case, the required machine resources may be relatively small, but the response times and swap counts are high. The statistics shown above result from lumping both types of transactions into the same class. Unfortunately, RMF does not gather enough data to directly separate the two types. For our purposes, however, it is safe to assume that most of the machine resources are consumed by transactions of the retrieval type.

Class BATCH has a similar interpretation problem. Most of the batch jobs appearing in this data are short (under one minute). However, most of the machine resources appear to be consumed by jobs that are much longer, in the 2-20 minute range.

Class CICS represents the CICS subsystem, which we understand is used mainly for data entry. In any event, the resources consumed by CICS are minuscule. The number of transactions is not available to RMF, since CICS handles its transactions internally.

Class MISC includes all RMF groups not described above. We can not tell exactly what sorts of activity are involved, but it's very unlikely to be associated with ARPS. The number of transactions in the MISC class is irrelevant.

To determine what fraction of the above usage is due to ARPS, we must rely on data obtained from analyzing the query patterns. That analysis is described in detail in Section 6.7. For our purposes here, it is enough to know that a two-week sample of interactive queries had the following characteristics:

Ground database:	275 queries at 67 average total seconds
	313 queries at 57 average total seconds
Aviation database:	384 queries at 21 average total seconds
	194 queries at 28 average total seconds

We also know from experimentation that the CPU resources for each query is approximately 50 percent of the total time, and that each query has about 5 CPU seconds startup, independent of the query size. In total, these queries consumed 30,711 CPU seconds total or 15,355 CPU seconds per week. The actual figure is perhaps somewhat higher, since we have no information on the queries against the Drug and Alcohol Database and we ignored the queries against the FECA, Flying Hours and Exposure Databases (approximately 8.5 percent of the queries in the evaluation set).

We know from the RMF data that the system averaged 34 percent CPU utilization during the sample periods, and that moderate and long TSO transactions comprise 51 percent of the load. If we assume that the system was that busy 8 hours per day, 5 days per week, then moderate and long TSO transactions consumed:

$$8 \times 5 \times 3600 \times 0.51 \times 0.34 = 24,970$$

seconds per week. This assumption implies that ARPS is directly responsible for 15,355/24,970 or 61 percent of the TSO load. Using 6 busy hours per day instead of 8 changes the estimate to 82 percent of TSO load. Either way, it is clear that ARPS queries are responsible for the bulk of the TSO load. We presume the same holds for batch.

6.1.3.2. I/O Capability

To estimate the impact of converting the ARPS database to some indexed form, we need an indication of how much disk I/O the ASMIS system can support. For each 15-minute interval, RMF acquires basically two numbers for each disk: total number of requests and total device busy time. These numbers are then converted to a total rate for the entire system and to the average service time (per request) for each disk.

The average service time for an unloaded disk appears to be around 0.025 seconds, which is consistent with published specifications. This number establishes an absolute maximum of 40 accesses per second that one could expect to satisfy by an individual disk. (The highest 15-minute average rate actually observed by RMF was 23 requests per second. The primary reason for this low figure is that there was not enough load on the system to saturate a single disk for an entire 15-minute interval.)

The total system I/O rate depends on channel and controller contention, as well as disk and cpu saturation. Across all intervals, the highest 15-minute average rate observed by RMF was 76 disk requests per second. It seems very likely that the system was I/O bound during this period. The cpu was only 43 percent saturated, although there were 20 TSO users and an average of 1.6 batch jobs. There was heavy activity on six disks and two tape drives, with substantial utilization of both primary and alternate disk channels. The next highest figure was 72 disk requests per second, again with 1.3 batch jobs and 15 TSO users, and only 54 percent cpu utilization. These observations suggest that the maximum achievable rate is around 80 disk accesses per second for the entire system.

6.2. SELECTING A DATABASE MODEL

In Section 6.1 it was determined that the IBM 4381 CPU was 34 percent utilized and the disk channels were 27 percent utilized. With this level of utilization, the system has adequate capacity to consider replacing the ARPS VSAM file structure with a database management system (DBMS). Before looking at the available DBMS products, a database model had to be chosen which would best provide the functionality to support the desired computer data environment.

There are three classes of data storage environments. There are files, transaction-oriented databases, and information systems. A list of these three data environments and a description of each is provided below:

6.2.1. Class I Environment: Files

In a file environment a database management system is not used. Files of data are generally not shared by applications, but are designed by the analysts and programmers when the application is created.

Characteristics:

- Applications start out to be simple, which initially makes them easy to implement. However, many times a large proliferation of files grow with high redundancy leading to high maintenance costs.
- Seemingly trivial changes to applications trigger a chain reaction of other changes and hence change becomes slow, expensive and tends to be resisted.

6.2.2. Class II Environment: Transaction-oriented Databases

A DBMS is used but without the degree of sharing found in an information system. Databases are production oriented where fast response time and high transaction rates (many records inserted, modified, and deleted) are important.

Characteristics:

- Thorough data analysis and modeling is needed, which takes time.
- Relationships between data must be predetermined.
- Lower maintenance costs than Class I.
- This environment leads eventually (but not immediately) to faster application development and direct user interaction with the database.

6.2.3. Class III Environment: Information systems

Databases are organized for searching and fast information retrieval rather than for high-volume production runs and provide good end user query facilities.

Characteristics:

Information system databases provide high flexibility which make them easy to implement and maintain.

6.2.4. The Current ARPS Data Storage Environment

ARPS currently belongs to the Class I environment. It is a file structure which by nature is difficult to maintain. According to the functional specification, however, ARPS should be in the Class III environment, an information system. As in ARPS, users of this type of system require ad hoc capabilities, fast retrieval, and data flexibility, all with code which is easy to maintain.

The database model which is chosen for a system depends primarily on the required data environment. The relational database model provides the capabilities needed in an information system environment. The relational approach has advantages over other database structures (such as the hierarchical or network models) because it provides:

1) Ease of use. Individual data items are grouped into records, which in turn are collected into tables. Use of these tables is intuitive even for staff members who are not computer experts.

2) Flexibility. Tables can be easily created and modified. Complex data structures can be represented.

3) Data independence. The data location and access methods are hidden from the application programmer. Only the DBMS software knows how the data is physically stored. This storage can change and the database tuned without users or applications being aware of it. The database software provides independence between tables and programs so that often existing programs do not require modification when the tables are revised.

4) Ad hoc query capabilities. Tables can be related as needed without having to predetermine likely combinations.

6.3. DATABASE EVALUATION

After the appropriate data environment was determined for ARPS and the relational database model was chosen, a list of possible DBMS candidates was compiled. The only criteria used for selection at this point was that the DBMS be relational and have the ability to run on an IBM 4381 under the MVS/SP operating system.

The initial list of products included:

DBMS	Vendor
DB2	IBM Corporation
Oracle	Oracle Corporation
IDMS/R	Cullinet
DATACOM/DB	Applied Data Research, Inc
SUPRA	Cincom

It is important to point out that rapid changes are occurring in the relational database market at this time. Since IBM announced its relational DBMS, DB2, there has been a great deal of activity by other vendors attempting to obtain a share of the IBM mainframe DBMS market. IBM currently has 50 percent of this market, with the remaining 50 percent distributed among the other vendors. IBM, Oracle, and Applied Data Research are among the vendors which have new versions of their relational databases scheduled for release in the fourth quarter of this year. These vendors have promised greatly improved performance, but the actual figures are not available at this time. This analysis does not attempt to take these promises into account, but rather is based only on vendor literature, interviews with users, and independent reviews as of July, 1988.

Several factors were examined in an effort to reduce the list of potential DBMSs to a manageable list for detailed study. These factors included the qualifications of the vendor, the probable future direction of the DBMS, and the ability of the DBMS to run effectively on the current hardware, operating system and telecommunications software. Running effectively meant that the DBMS must be able to respond to queries in a similar amount of time as the current system for the number of simultaneous users accessing ARPS, currently averaging 20. With these criteria in mind, three DBMS systems were eliminated from further consideration including, IBM's DB2, Oracle Corporation's Oracle, and Cullinet's IDMS/R.

6.3.1. DB2

DB2 was removed from the list based on its performance on an IBM 4381 under the MVS/SP operating system and the fact that it is unlikely that IBM will attempt to improve the performance of DB2 in that configuration.

IBM plans to release a new version of DB2 in October 1988, but this version will not run on the MVS/SP operating system. It runs only on MVS/XA or the newest version of MVS, MVS/ESA. Many of the performance improvements claimed under the new version of DB2 can only be realized under MVS/ESA, and only the E Series of mainframes can run ESA. The Army Safety Center's IBM 4381 MG13 would have to undergo a major hardware upgrade in order to run ESA. This upgrade would be to the model group 91E or 92E.

As evidence of DB2's latest release, IBM is focusing its DB2 performance improvements in the MVS/ESA operating system. As discussed in (Bucken, 1988), those sites which are committed to DB2 and do not have a large E Series mainframe will need to have hardware upgrades to run ESA in order to fully gain the performance enhancements that the new versions of DB2 will have in the future. Furthermore, DB2 has been found to be too resource intensive for the IBM 4381, and its use is not recommended on that machine (Schussel, 1986).

6.3.2. Oracle

Oracle was eliminated from the list due to its performance and its inability to support even the average number of simultaneous users of ARPS under the MVS/SP operating system.

Oracle has just recently ported its relational database system to the MVS operating system, so there is very little information available about Oracle's performance in the MVS environment. The only third party performance information we obtained was a benchmark comparison by B. Monsanto (Babcock, 1987). The article stated that Oracle's performance was marginally worse than that of DB2 in all of its transaction processing tests. Benchmarks provided by Oracle showed that they outperformed DB2, but to maintain a constant performance advantage over DB2, Oracle had to be extremely well tuned.

It was also found that Oracle was unable to support the current number of users accessing ARPS. Under the current version of Oracle, Version 5.1, there is a limit to the number of simultaneous users that can be supported under the MVS/SP operating system. The number of users is a function of the amount of Private Area available on the computer system. System Analyst, Jim Hayes, estimated the Private Area available on the ASMIS system to be 5MB. This figure was used in the formula given by Oracle to determine the number of concurrent ORACLE users that can be supported on MVS/SP. The limit was found to be approximately 18 users. This is an unacceptable restriction on the number of users and would not be sufficient to support the peak periods of database access. Since Oracle is unable to support an adequate number of concurrent users and DB2 was eliminated based on similar performance, Oracle was eliminated from further consideration as well.

6.3.3. Cullinet

Cullinet's IDMS/R was eliminated from the list based on the anticipated future direction of the vendor. Cullinet is shifting away from its previous emphasis on providing a DBMS for IBM systems. Cullinet is believed to be primarily interested in providing database support products for other DBMS, such as DB2, that run on the IBM (Datapro, 1988). They seem to be shifting their emphasis to their newly acquired DBMS which runs on VAX systems. As a result of this, their commitment to supporting and developing their IBM DBMS in the future is in doubt.

6.4. DATACOM/DB VERSUS SUPRA

After eliminating the above DBMSs based on technical grounds or their uncertain future, the remaining database management systems were:

DBMS	Vendor
DATACOM/DB	Applied Data Research, Inc
SUPRA	Cincom

DATACOM/DB and SUPRA both incorporate the basic facilities of well integrated, fully functional database management systems. The following features are provided by both databases. (The sections discussing the corresponding requirements of ASMIS are listed in parentheses.)

1) Data Independence. The location of the data in the database and the access methods to it are hidden from the application programmer (Sections 4.2.6 and 4.3.6).

2) Data clustering capabilities. The database system provides the capability to cluster data so related data can be stored in the same physical block. This capability improves data retrieval performance by allowing related records to be accessed with a single IO event.

3) Database performance monitoring and tuning facilities. Facilities are provided to monitor and fine tune the performance of the DBMS (Section 4.2.5).

4) Data security and access control. Facilities are available to define security for the data and to allow access to portions of that data by individual users (Sections 4.4.1 and 4.2.3).

5) Integrated data dictionary. The data dictionary is a central library for defining all the data elements, fields, entities, synonyms, cross-references and the relationships between them. The data dictionary also maintains the user access control information for the database,

such as read, write and modify capabilities (Sections 4.2.2, 4.2.3, 4.2.6, 4.2.7 and 4.4.1).

6) Applications development flexibility. Application programs which access the database may be written in COBOL or a fourth generation language available with the DBMS. The fourth generation language provides greater productivity for the staff after they learn the new language. COBOL will still be needed for some applications because some operations are too complex for a fourth generation language to handle (Section 4.3.6).

7) Query language and report generator. Utilities which provide a easy-to-use ad hoc query facility and the generation of simple reports are available (Sections 4.3.3 and 4.2.2).

8) Transaction processing capabilities. When a transaction (a logical unit of work) is interrupted by a serious error, the entire transaction is automatically rolled back. This prevents the error from causing unwanted changes to the database and returns the data to its previous condition (Section 4.4.3).

9) Automatic restart capabilities. In case of system failure, automatic restart capabilities are provided to recover the database with minimum effort (Section 4.4.3).

10) Utility programs to create and maintain the database. Facilities are available to assist in data reorganization and data security. These types of programs and facilities aid in providing data structure flexibility (Sections 4.2.6 and 4.2.7).

11) Export and import utilities. Utilities are provided which allow moving data between flat files and the database (Sections 4.3.4 and 4.3.5).

It is our conclusion that both DATACOM/DB and SUPRA would be able to satisfy the requirements of ASMIS. There is however, a significant difference in the cost of the systems to the Army Safety Center. The United States Army has an agreement with Applied Data Research allowing any Army facility, including the Army Safety Center, to acquire DATACOM/DB and related products at no additional cost. SUPRA, on the other hand, would have to be purchased by the Army Safety Center at the GSA price of approximately $269,000, excluding installation and training. Although SUPRA is a quality product, it provides no features or facilities over DATACOM/DB that can justify purchasing it. For this reason alone, SUPRA was eliminated from further consideration.

There are several advantages gained by the Army Safety Center in using DATACOM/DB. A strong user community exists within the Army and many sites with the same hardware configuration are using DATACOM/DB. In addition to the database related software, there are other products available under the Applied Data Research contract such as LOOK, (measures system performance and identifies system parameter problems), ROSCOE (an online programming tool) and the LIBRARIAN (manages software development and maintenance), to name a few.

Below is a more detailed list of additional information and features that DATACOM/DB provides:

6.4.1. Software and Services Available

DATACOM/DB can be acquired by writing a letter requesting the Applied Data Research baseline products. This letter can be sent to the following address:

Commander United States Army,
Information Systems Software Center
ATTN: ASBI-CDE
Stop C70
Ft. Belvoir, VA 02260-5456

The list of available software under the original contract includes:

- DATACOM/DB - general-purpose, high-performance, relational DBMS.
- DATADICTIONARY - a repository for all information pertaining to data descriptions of files, records, dataviews, keys, and fields. It is also where information on jobs, systems, and programs can be stored. As a tool for the Database Administrator, it can assist in measuring and analyzing data usage throughout the system.
- DATAQUERY - end-user-oriented information retrieval and data manipulation facility which operated under CICS and can also function in batch mode.

- DATAREPORTER - a information retrieval and report generation system.
- VSAM Transparency - allows migration from VSAM applications to DATACOM/DB without any modification to application programs.
- MetaCOBOL - provides a high-level COBOL programming environment. It is an extension of COBOL which enforces structured programming.
- ADR/DL - high level COBOL preprocessor.
- ROSCOE - online programming tool.
- LIBRARIAN - manages software development and maintenance.
- LOOK - provides performance measurement in the IBM mainframe environment.

Software and services to be offered under the second contract to become available in October of 1988 include:

- ADLIB - allows programmers to use PCs to develop COBOL applications.
- DATASECURE - provides cryptographic security measures to DATACOM/DB databases. It is designed to support the federal government Data Encryption Algorithm (DEA).
- IDEAL - a fourth-generation language which greatly increases programmer productivity in developing online applications in the CICS environment.
- eMAIL-VOICE - electronic mail.
- ADR/ETC - a tool for extended text creation which can be delivered by the electronic mail facility, eMAIL.
- DATACOM/PC provides upload/download capabilities to/from the mainframe DATACOM/DB DBMS, extended reporting, and support for standard PC file formats that enable the user to operate with microcomputer programs such as Lotus 1-2-3, dBase III and Multiplan.
- ADR/D-NET - supports a distributed database environment.

The first thirty-five days of on-site installation and training is free. Additional training can be received at $1,000 per day plus travel and per diem for the instructor.

6.4.2. SQL

Version 8.0 of DATACOM/DB, available in September of 1988, delivers full support for the Structured Query Language (SQL). SQL is the ANSI standard query language for relational DBMS. Support for SQL establishes an open architecture that allows DATACOM/DB users to take advantage of applications and tools offered by other vendors, while protecting their investment in existing applications. It allows applications developed for other SQL-based relational systems to run on DATACOM/DB and vice versa, achieving application portability.

Full support for SQL is embedded in the DATACOM/DB nucleus, making it an integral part of DATACOM/DB's relational architecture. The ADR software products which will support DATACOM/DB's SQL will include:

- IDEAL
- DATAQUERY
- Static and dynamic embedded SQL - ADR/DL, COBOL

6.4.3. PC to Mainframe Support

DATACOM/PC provides the functions of ADR/Dataquery on an IBM PC. In addition DATACOM/PC features upload/download facilities, extended reporting, and support for standard PC file formats that enable the user to operate with microcomputer programs such as Lotus 1-2-3, dBase III and Multiplan. Data transfer to other PC-based programs can be accommodated using DIF (data interchange format).

6.4.4. Variable Length Strings for Narrative Data

DATACOM/DB records are defined as fixed length records, but compression techniques may be used to free unused space in sparsely filled text fields.

6.4.5. Performance

According to a benchmark performed by William Inmon (Inmon, 1988), DATACOM/DB had the best performance statistics in both the database requests per second and the transactions per second of all the database systems considered in his analysis.

6.5. DATACOM/DB TRAINING

ADR provides courses covering the tools and techniques necessary to use their DBMS software. Courses for applications programmers, database designers, performance monitors and systems programmers in charge of their local MVS systems are available. Because no definite implementation for the ASMIS system has yet been developed, the exact courses necessary for the USASC ADP staff is not completely known. The paragraphs below identify the ADR courses which could be necessary for personnel performing various functions in the enhanced ASMIS system. The exact courses necessary can not be determined until a detailed design of the enhanced ASMIS system including more information on the types of DBMS software to be used is completed. This information would then guide the training of the USASC ADP staff.

For applications programmers a number of courses could be required. To provide the basic information necessary to use the DBMS, each applications programmer should attend DB100, DATACOM/DB Application Programming and LI101, The Librarian. For the programmers who will develop applications using IDEAL, a fourth generation language the course ID010 IDEAL CBT Series would be necessary. For the programmers who will develop applications using COBOL under CICS, DB251, DATACOM/CICS Service Facility and DL101, ADR/DL Application Programming for DATACOM/DB would be necessary.

For the design and implementation of the database, MT102, Conceptual Data Modeling and MT251, Datacom/DB2 Database Design should be taken by the Database Administrator and the DBA backup.

For monitoring the performance and fine tuning the DBMS and its associated utility programs, four courses are available. For overall monitoring of the DBMS, LO251, Using the LOOK CICS Monitor and LO151, LOOK MVS Usage should be taken. For ADR/IDEAL, a fourth generation language, ID251, IDEAL Site Administration and DB252, Methods of Tuning the DATACOM/IDEAL Environment should be taken. These courses could be taken by the Database Administrator and the DBA backup. Or this load might be shared between the Database Administrator and the DBA backup and the programmer in charge of maintenance of the MVS system.

For installation and support of the ADR system on the IBM 4381, the course DB151, Establishing and Supporting the DATACOM Environment should be taken by the system maintenance programmer.

6.6. MVS/SP VERSUS MVS/XA

Currently the Army Safety Center's IBM 4381 is running the MVS/SP operating system. However, there are two newer versions of MVS available, MVS/XA and MVS/ESA. Since 1983, IBM has made no significant extensions to MVS/SP, but instead has concentrated its efforts on MVS/XA and its successor, MVS/ESA. Unfortunately, MVS/ESA only runs on the E Series of mainframes. The Army Safety Center's IBM 4381 MG13 would have to undergo a major hardware upgrade to the model group 91E or 92E in order to run MVS/ESA. However, an upgrade to MVS/XA under the current hardware configuration is possible.

With MVS/SP there is a 16 MB restriction of virtual memory for each user. Jim Hayes has expressed a concern that this limit may be reached in the near future with the statistical software package, SAS. As new capabilities have been added or changes made to SAS, the software has required more address space in the private area in order to run. MVS/XA extends this addressing limit to 2 Gigabytes. In addition to extending the virtual memory, MVS/XA also provides I/O performance improvements.

Table 6.1 shows the IBM software products which would need to be purchased in order to upgrade to MVS/XA. Note that there are two purchase options, a one time charge or 36 monthly payments. There are no additional charges (such as set up fees) encountered by choosing the 36 monthly payments. Table 6.2 shows the MVS/SP products currently running on the ASMIS IBM computer system and their monthly charges. These charges would be replaced by the upgrade to MVS/XA. The additional monthly charge that USASC would have to pay for 36 months in order to upgrade to MVS/XA is $3,151. However, it is important to note that after 36 months the MVS/XA software would be paid in full whereas under the current leasing agreement for the MVS/SP products, those charges will continue for as long as ASMIS uses the software.

TABLE 6.1. Options for New Product Purchases Required by MVS/XA

Program #	Description	One Time Charge	Monthly Charge	Term
5665-XA2	MVS/XA Data Product Facility	31050.00	1027.00	36 mo
5665-274	RMF Version 3	20718.00	685.00	36 mo
5665-285	TSO/E (MVS/370) (MVS/XA)	14320.00	473.00	36 mo
5740.XC6	MVS/SP-JES2 VER 2	126116.00	4174.00	36 mo

	Total One Time Charges	192204.00		

	Total FLPP Charges		6359.00	

TABLE 6.2. MVS/SP Products Displaced by Upgrading to MVS/XA

Program #	Description	Monthly Charge
5740-XY4	RMF Version 2	456.00
5740-XYS	MVS/SP-JES2 R3.6	2220.00

	Total Program Product Removals	2676.00
	Total Additional Monthly Charge	3683.00

Upgrading to MVS/XA would allow ASMIS to take advantage of current MVS enhancements, and eliminate the possibility of encountering the address space limitation under MVS/SP.

6.7. ARPS QUERIES

The ARPS program was modified so that all messages going to the user at his terminal and all his responses were saved. At the end of the user's session, this information was saved. This resulted in trapping all queries for all users who did not let their terminals time out. For this evaluation, the queries from the Aviation, Ground, FECA, Exposure and Flying Hours options of ARPS were logged. Approximately two weeks of logging was provided for this evaluation. Because each database is queried by a unique version of the ARPS program, implementation of the code to enable logging occurred over 3 days, May 11, 12 and 13. The final day for logging was May 27. The distribution of queries in this evaluation set is shown in Table 6.3.

TABLE 6.3. Distribution of Queries in the Evaluation Set

Database	Number of Queries	Percent of Queries
Aviation	591	42.3
Ground	689	49.3
FECA	90	6.5
Flying Hours	18	1.3
Exposure	9	.6

To evaluate the appropriateness of a relational database management system for the ASMIS system, the following steps were taken for each database:

1) Description of the evaluation set. The set of sample queries was characterized. Included here is information on number of matrix versus non-matrix queries and the number of database cases required by each query.

   The log of queries did not include the number of records retrieved from the database. To obtain a count of the number of database cases that were needed to generate the display described in the query, the queries were divided into two types, those that produced a matrix and those that did not. The first type, those queries that do not produce a matrix, create a columnar display of the requested fields. The log of each of these queries includes the line "YOU HAVE DISPLAYED xxx LINES - DO YOU WISH A PRINTED REPORT? (Y)ES (N)O (S)TOP." This message was used to estimate the number of lines displayed for the user. Since the number of lines displayed is the same as the number of database cases, this number serves as an estimate for the number of database cases retrieved. Since some data display can occur after the final message, 50 was added to this number. If no such message was included, 50 lines was used.

   For the queries which produced a matrix, no information in the query log could be used to estimate the number of cases which would need to be retrieved. Instead, a random sample of the queries generating a matrix were used to estimate the number of cases which would need to be retrieved.

2) ARPS performance. A model for the current ARPS program was developed and the sample queries were evaluated using this model.

3) Organization of data fields for a relational DBMS. Prior to modelling the relational database system, the current ASMIS data fields were organized to reflect an appropriate structure for a relational DBMS. This organization is NOT the final organization, rather it is a first approximation. As will be seen, the initial relational organization fails to account for some important features of the access style of the current users.

4) Expected DBMS Performance. A model for accessing the data using a relational DBMS was developed and the sample queries were evaluated using this model.

5) Analysis of the DBMS Performance. The DBMS performance is compared to the ARPS performance, changes to the proposed DBMS structure are made and the DBMS performance is reevaluated. The query patterns of the individual users are evaluated and anomalies identified. For the anomalies, more appropriate user behavior is supplied and the DBMS model is again evaluated and the performance compared to ARPS.

6.7.1. Ground Database

6.7.1.1. Description of the Evaluation Set

Six hundred eighty nine (689) of the analyzed queries used the ground database. Of these, 376 (55%) produced the non-matrix style output, 313 (45%) produce the matrix style output. Of the non-matrix style output, 101 jobs will not be included in the following analysis. Of these, 81 jobs were routed to the batch queue by the user, 7 jobs involved PROC statements, 9 jobs were CASE PRINTS and 4 jobs had miscellaneous errors which we believe means that these jobs were not run at all. In this analysis, the 81 batch jobs will be excluded because we are characterizing the interactive response of the proposed system, not its batch characteristics. The remaining 275 non-matrix jobs will be used to characterize one aspect of the interactive database access required for the ground database. The 313 matrix jobs will be used to characterize the other aspect of the database access.

The 275 non-matrix jobs displayed an average of 258 cases. Looking at these cases, revealed two subgroups (see Table 6.4). Two-hundred sixty (260) of these jobs displayed all the database cases that were chosen by the selection criteria. This group of queries is characterized by a large number of short requests and a few large requests. Seventy-five (75) percent of these jobs displayed 50 cases or less; 84 percent displayed 150 cases or less; 95 percent displayed 950 cases or less. The average number of cases for this group of jobs is 206.

The second group of non-matrix jobs consists of those that were terminated by the user after displaying only part of the database cases meeting the selection criteria (the user answered S to the question "YOU HAVE DISPLAYED xxx LINES - DO YOU WISH A PRINTED REPORT? (Y)ES (N)O (S)TOP"). The average number of cases displayed by this group is 1163. Examination of these 15 queries revealed no obvious reason for the distinction.

The characterization of the matrix jobs was done by selecting a random sample and running the chosen jobs to obtain the number of database cases used to produce the matrix. Twenty queries was chosen as a workable sample. The selection was done by taking every 13th matrix query, resulting in 22 matrix jobs. Of these, twenty resulted in database cases being selected. For these jobs an average of 1521 database cases were read. The two jobs that selected zero records had bad selection criteria (included both TYPE = B and TYPE = D). Table 6.5 includes the counts of database cases retrieved for each of the 22 matrix jobs run.

TABLE 6.4. Estimated Number Cases Displayed for Non-Matrix Queries Done as Interactive Jobs

Completed Jobs		Jobs Stopped by user	
Number of Queries	Number of Cases	Number of Queries	Number of Cases
196	50	6	150
24	150	3	250
13	250	1	450
4	450	2	650
3	550	1	3950
2	650	1	4350
3	750	1	5750
3	850		
1	950		
1	1150		
2	1250		
1	1350		
3	1650		
1	1750		
1	2150		
1	6150		
1	6450		

Completed jobs: 260, average of 206 cases displayed
Jobs stopped by user: 15, average of 1163 cases displayed
Total jobs: 275, average of 258 cases displayed

TABLE 6.5. Number of Database Cases Read for a Random Sample of Queries Generating Matrices

Number of Queries	Number of Cases	Number of Queries	Number of Cases
1	0	1	672
1	0	1	694
1	11	1	773
1	212	1	819
1	218	1	1286
1	218	1	1383
1	218	1	1737
1	252	1	2816
1	254	1	3762
1	351	1	3800
1	638	1	10301

20 jobs, average of 1521 database cases read

6.7.1.2. ARPS Performance

The log data we received contained no information about the amount of time required to complete a query using the current ARPS program. To provide a standard for comparison, a performance model for the current ARPS program was derived. The time derived is based on the wall clock time to do a query on an otherwise idle system. Thus the time includes both CPU and IO times.

The assumptions for the model are:

- Cases to be searched are specified by the start and end dates included in the query specification.
- Database cases are uniformly distributed over time.
- No cost for formatting the output for the user is included in the time estimates. Checking the cost for formatting the non-matrix output revealed that the output formatting added less than 0.5 percent on average to the calculated times.

The model does the following operations for each case:

1) Read the record.

2) If the personnel or equipment data is needed, expand the personnel and equipment data. Our test indicate very little difference between accessing personnel data, equipment data, or personnel and equipment data. Thus our model assumes expansion of the whole case if either personnel or equipment fields are used.

3) Evaluate the selection criteria to determine if the record is selected or not.

Thus there are two types of retrievals done within the current ARPS, those that use only the basic record and those that use the basic data plus personnel and/or equipment data. To obtain numbers to use with this model, sample runs were made. Each sample run retrieved all the records from January 1, 1981 to the present and produced a matrix which included the total number of records retrieved. The record count plus the wall clock time for the retrieval was used to determine the number of records per second which could be read and processed. Reading only the basic record took 2 minutes 45 seconds and returned 141,290 records. Reading the record and using either the personnel and/or equipment data, took 4 minutes and 27 seconds. The formula used to predict the time required to respond to a single query is:

time = number-days * number-record/day * time-to-do-1-record

where:

number-days	is the number of days to be searched, (from the start and end date specified in the query).
number-records/day	is the average number of records per day, (average based on number of records in 1981 through 1986).
time-to-do-1-record	is obtained from the sample runs described above.

Using this formula and the same queries used to evaluate the DBMS implementation, performance for the current ARPS program was derived. Table 6.6 is the performance for the queries which produce non-matrix output. Table 6.7 is the performance for the queries which produce matrix output.

TABLE 6.6. Time to Complete Non-Matrix Queries Using Model of Current ARPS Program

No. of Queries	No. of Seconds	Cumulative Percent	No. of Queries	No. of Seconds	Cumulative Percent
10	0.07	4	1	39.68	64
26	0.11	15	2	48.22	64
1	0.21	15	2	49.57	65
4	0.32	17	4	61.77	67
1	0.69	17	1	63.52	67
1	1.71	17	6	65.17	70
1	1.98	18	1	72.41	70
3	2.05	19	1	73.5	70
1	2.46	19	1	74.29	71
2	3.32	20	17	78.03	78
1	3.47	21	1	100.96	78
1	3.69	21	1	119.25	79
1	4.43	21	1	120.6	79
1	4.58	22	1	120.66	79
1	4.64	22	1	123.58	80
1	5.61	23	1	128.73	80
1	5.97	23	7	141.17	83
1	6.07	23	1	141.28	83
1	6.57	24	2	156.11	84
1	9.07	24	2	156.21	85
2	9.82	25	2	169.23	86
1	13.01	26	2	178.88	87
3	15.15	27	1	180.53	87
1	15.2	27	3	195.15	88
1	15.79	28	10	195.25	92
1	19.22	28	1	195.3	93
1	19.63	28	1	208.31	93
1	22.83	29	1	219.35	94
32	24.1	42	1	219.78	94
1	24.16	42	1	248.63	94
4	31.1	44	2	250.45	95
1	33.42	44	1	297.32	96
1	33.75	45	1	298.28	96
1	36	45	6	298.71	98
1	38.17	45	1	298.82	99
42	38.99	62	2	322.13	100
2	39.1	63	1	327.67	100

Average time: 67 seconds

TABLE 6.7. Time to Complete Matrix Queries Using Model of Current ARPS Program

No. of Queries	No. of Seconds	Cumulative Percent
7	9.82	35
1	24.1	40
1	24.59	45
5	38.99	70
1	65.17	75
1	117.07	80
1	120.6	85
1	141.28	90
2	195.15	100

Average time: 57 seconds

6.7.1.3. Organization of Fields for Relational DBMS

The ground database was analyzed to determine an appropriate relational structure. The current database fields which relate to a single concept, for instance an individual involved in an accident, are grouped into a single relation. This structure is only a first approximation, to be used for modelling the DBMS performance. It is not necessarily the appropriate structure for implementing the database. The relations identified are:

- Basic data. The description of the accident and summary information such as total damage cost, total cost, etc.
- Personnel data. The description of one person's involvement in the accident. There may be multiple personnel records for each accident.
- Equipment data. The description of a single piece of equipment and its involvement in the accident. There may be multiple equipment records for each accident.
- Environmental data. The description of the environment at the time of the accident. There may be multiple environmental records for each accident.

- Cause data. The description of the causes of the accident. There may be multiple cause records for each person involved in the accident. This includes the task(s) assigned to the individual and the error(s) associated with those tasks.
- Narrative data. This is the whole narrative for the accident. It includes a description of the accident, the causes, the corrective actions, etc.
- UIC translation data. This is the method of translating from the UIC structure at the time of the accident to the current UIC structure. In the current implementation it includes the fields phantom UIC, phantom station, etc.

To determine how the evaluation set of queries would be affected by this restructuring of the database, each field was assigned to one of the relations identified above and the full set of 689 queries was scanned. For each query, the number of relations referenced in the selection criteria, the number of relations referenced in the fields to be displayed and the list of which relations were accessed were tabulated. This information was used to project the performance of the DBMS model. Appendix C.1.1 is the list of fields used in the selection criteria and the associated frequency of use. Appendix C.1.2 is this same list, ordered by frequency of use and thus forms the basis for the database keys for the relations in the ground database. Appendix C.1.3 is the list of fields used in the data display.

6.7.1.4. Expected Database Performance

A database management system uses indexes to speed access to the data. An index for Field A is a data structure within the DBMS which lists all records having each value for the Field A. For instance, an index for the ONPOST field in the ground database would have two values, A (on post), B (off post), and each database case would be associated with one of these two values. To find all accidents which had been marked as "on post," the DBMS would use the ONPOST index to identify the database case identifiers and then obtain these cases directly. If the query involved two selection criteria, [for instance, ONPOST = A and STATION = 01767 (Ft. Rucker, AL)], the DBMS would get the list of database case identifiers for ONPOST = A, and the list of identifiers for STATION = 01767, and pick out the identifiers which were in both lists. This should result in a relatively small list of database cases to be retrieved. Because relatively few cases will be retrieved, the probability of getting two of the selected cases from the same disk access is small. Thus, for estimation purposes, each database case selected will require a disk access.

In contrast to the DBMS, the current ARPS program would find all these cases by reading the whole ground database file and checking the ONPOST and STATION fields in each case. Because the ARPS program reads the ground database sequentially, it gets multiple cases per disk access.

In considering replacement of the current ARPS program with a DBMS, an important question is: Can the DBMS retrieve few enough records that its performance on a large collection of queries will be as good as or better than the performance of the current ARPS program? Phrased another way, averaged over a large number of queries, can the DBMS which retrieves fewer records per second (remember one disk access gets only one case), but needs to retrieve fewer records, beat the current ARPS program which retrieves and examines all cases?

To evaluate this question, a simple model for a relational DBMS was created. This model reflect only the I/O time required to retrieve the data. It presupposes no knowledge of the actual physical location of the data (i.e., how the location of the basic data for a case is related to the location of the personnel data for the same case). Thus, an estimate of the time to perform each query was based on the time to access the necessary data from the disk. The following formula was used:

$$\text{time} = \frac{\text{\#cases} * \text{\#total-relations-used} * \text{\#disk-accesses-per-relation}}{\text{\#disk-accesses-per-second}}$$

where:

#cases   is the number of database cases retrieved by the query. This number comes from each query in the evaluation set.

#total-relations-used is the total number of relations used by the query. This includes relations used in the selection criteria and in the fields to be displayed and was computed for each query in the evaluation set.

#disk-accesses-per-relation to locate a specific database case requires multiple disk accesses. The index on database case identifier must be searched to locate the specific database case. For this estimate, 3 was chosen. This represents the average number of disk accesses to retrieve a specific database case from a DBMS containing 40,000 to 8,000,000 records.

#disk-access-per-second is the number of disk accesses the disk drive can do in one second. This was chosen to be 40 and is based on the maximum possible for the disk drives and controller currently used.

For the non-matrix queries, the average time was 52 seconds; Table 6.8 shows the distribution. For the matrix queries, the average time for a sample of 20 queries was 213 seconds; Table 6.9 shows the distribution.

TABLE 6.8. Estimated Time to Complete Non-Matrix Queries Using the DBMS Model

No. of Queries	I/O Time No. of Seconds	Cumulative Percent	No. of Queries	I/O Time No. of Seconds	Cumulative Percent
63	3.75	26	1	187.5	94
48	7.5	45	1	191.25	94
62	11.25	70	2	195	95
13	15	75	1	225	95
4	22.5	77	1	247.5	96
12	33.75	82	2	255	96
5	37.5	84	1	258.75	97
7	45	87	1	281.25	97
7	56.25	89	1	285	98
1	75	90	1	371.25	98
2	82.5	91	1	405	98
1	112.5	91	1	978.75	99
1	123.75	91	1	1293.75	99
3	135	93	1	1451.25	100
1	146.25	93	1	1845	100

Average time: 52 seconds

TABLE 6.9. Estimated Time to Complete Matrix Queries Using the DBMS Model

No. of Queries	I/O Time No. of Seconds	Cumulative Percent	No. of Queries	I/O Time No. of Seconds	Cumulative Percent
1	1.65	5	1	115.95	60
1	31.8	10	1	122.85	65
3	32.7	25	1	130.275	70
1	37.8	30	1	192.9	75
1	38.1	35	1	282.15	80
1	52.65	40	1	311.175	85
1	95.7	45	1	422.4	90
1	100.8	50	1	570	95
1	104.1	55	1	1545.15	100

Average time: 213 seconds

6.7.1.5. Analysis of Database Performance

Comparison of the ARPS performance and the projected DBMS performance shows little difference in the time required to handle the non-matrix queries: 67 seconds for ARPS and 52 seconds for the database (see Tables 6.6 and 6.8). Comparison of Tables 6.7 and 6.9 shows a definite difference in the handling of matrix queries. ARPS processes the 20 queries in the random sample in 57 seconds. The projected database time is 213 seconds, or 3.7 times longer. For purposes of the comparison below, the non-matrix times will be judged to be equal. The non-matrix queries are 55 percent and matrix queries are 45 percent of the total ground database queries. Applying the following calculation and considering the load caused by ARPS to be 1.0, the load caused by the projected database implementation is then 2.2 or more than twice the ARPS load.

%-non-matrix * speed-factor + %-matrix * speed-factor = DBMS load

.55 * 1.0 + .45 * 3.7 = 2.2

The method of prediction used in Section 6.7.1.4 specified that the location of the database cases could be done entirely with indexes into the database. This is true in the organization described in Ssection 6.7.1.3 only if the query needs arguments from only one of the relations listed. This is true in 57 percent of the queries in the evaluation set. Thus, given the database structure proposed in Section 6.7.1.3, 43 percent of the queries would require more time. For the purposes of this estimate, this problem can be alleviated by restructuring the database. To allow 95 percent of the queries to be done entirely with keys, requires two new relations. These two relations are a combination of the basic and personnel data and the basic and equipment data. The database now has the following relations:

Basic data
Personnel data
Equipment data
Environmental data
Cause data
Corrective Action data
Narrative data
Basic-Personnel data
Basic-Equipment data

Note that this would NOT be the method used to implement the database. This is for estimation only and serves to identify structure within the ground database which must be more fully studied and characterized before the ground data is placed in a relational DBMS. More careful characterization of these queries will identify the underlying structure.

The remaining five percent of the queries would still take longer than predicted in Section 6.7.1.4. Of these 1.1 percent are narrative searches. The selection in these jobs was on fields in the basic data plus the search of the narrative. Approximately three percent (3.2%) specified an argument which involved fields from the basic, personnel and equipment relations. The effect of these new relations on a query with a selection criteria using the basic, personnel and equipment relations is unknown but it would be an improvement.

The database design described in Section 6.7.1.3 specified no physical storage constraints. In particular, there is nothing known about the relationship between the location of the basic data and the personnel or equipment data for a case. This implies that queries that get data from more than one relation require more time than queries which can be totally satisfied by one relation. By providing constraints on the physical location for the data for a case, we can reduce the time required. Recomputing the times for all queries to reflect the change in the physical storage results in the matrix jobs taking an average of 114 seconds and the non-matrix jobs taking an average of 17 seconds. Redoing the calculation above gives:

%-non-matrix * speed-factor + %-matrix * speed-factor = DBMS load

.55 * 0.25 + .45 * 2 = 1.04

The time for the DBMS to do matrix queries is still twice that of ARPS. Anything that would reduce the number of matrix queries, would greatly improve the overall performance of the DBMS-based system. The queries were again analyzed, this time looking for users who made a large number of queries during a session. The eight sessions with ten or more queries per session were analyzed. The 12 queries done by SCSS55 were non-matrix queries retrieving a small number of cases each and thus of little interest. In the 11 queries done by SCPAE3 there was no pattern to the selection criteria. The remaining six sessions were 136 matrix queries and there was a pattern to the selection criteria used in the whole session. In all six cases, there was a major selection criteria which affected all of the queries done. In each session the user also specified another selection criteria which affected some but not all queries. For example, user EUHQAA specified TYPE = A in each of his 10 queries. For part of these queries, this was the whole selection criteria. For the remainder of his queries, he added either ACTIVCAT = 16 or ACTIVCAT = 28. Basically these six people were using ARPS to produce a number of statistical tables based on the same selection criteria. This work could be done more efficiently in a package specifically suited for the generation of statistical tables.

Replacing the 136 queries attributed to the six sessions described above with six queries, results in a reduction in the total number of matrix queries to 183. This results in reducing the matrix query load to 58 percent of the original. Applying this to the load formula above results in:

%-non-matrix * speed-factor + %-matrix * speed-factor = DBMS load

.55 * 0.25 + .45 * 2 *.58 = 0.66

The above DBMS model is based on I/O time to obtain the requested records. In the aggregate, this model allows as much CPU time as I/O time because when multiple queries are running simultaneously, one query is using the CPU while others are using the I/O subsystem. The projected DBMS load indicates that the DBMS-based system could use 1.5 times as much CPU as I/O and still take no longer than the current ARPS system.

Because of the relative large role of repeated selection of database cases for the matrix queries selected above, the whole set of queries were analyzed to see what the role of previously selected data might be. At the end of each query, the user is asked what he wishes to do next and if appropriate whether he wishes to do further displays based on the data already selected. Two hundred forty (240) queries (of a total 689) had an already selected set and the user indicated that the data should be reused. This means that 35 percent of the all queries could be satisfied using data already selected by the DBMS. This does not imply that the already selected data can necessarily be used directly. The current ARPS implementation filters these cases again using the selection criteria the user specifies. Thus, the user is free to further restrict the selection.

Looking more closely at the set of queries which had an already selected set shows these 240 "reuse" responses were 62 percent of the queries where the "do further displays based on the data already selected" was appropriate. Reusing already selected data is appropriate, if you are not setting up a batch job and if this is not the last query in your session. Thus, 62 percent of the time the user could have used the already selected data subset. This clearly indicates that there is a pattern to the queries submitted by a single person and thus the new implementation of ARPS should use this knowledge to reduce the load on the DBMS.

6.7.2. Aviation Database

6.7.2.1. Description of the Evaluation Set

Five hundred ninety one (591) of the analyzed queries used the aviation database. Of these, 397 (67%) produced the non-matrix style output, 194 (33%) produced the matrix style output. Of the non-matrix style output, 13 jobs will not be included in the following analysis because they were routed to the batch queue by the user. The remaining 384 non-matrix jobs will be used to characterize one aspect of the interactive database access required for the aviation database. The 194 matrix jobs will be used to characterize the other aspect of the database access.

The 384 non-matrix jobs displayed an average of 128 cases. Looking at these cases, revealed two subgroups (see Table 6.10). Three hundred seventy one (371) of these jobs displayed all the database cases that were selected by the arguments specified. This group of queries is characterized by a large number of short requests and a few large requests. Eighty-two (82) percent of these jobs displayed 50 cases or less; 88 percent displayed 150 cases or less; 98 percent displayed 950 cases or less. The average number of cases for this group of jobs is 123. The second group of non-matrix jobs consists of those that were terminated by the user after displaying only part of the

TABLE 6.10. Estimated Number Cases Displayed for Non-Matrix Queries Done as Interactive Jobs

Completed Jobs: Number of Queries	Completed Jobs: Number of Cases	Jobs Stopped by User: Number of Queries	Jobs Stopped by User: Number of Cases
306	50	9	150
22	150	1	250
19	250	1	450
5	350	1	550
6	450	1	850
1	550		
1	750		
3	850		
1	1050		
2	1150		
2	1350		
1	1850		
1	2950		
1	3050		

Completed jobs: 371, average of 123 cases displayed
Jobs stopped by user: 13, average of 265 case displayed
Total jobs: 384, average of 128 cases displayed

TABLE 6.11. Number of Database Cases Read for a Random Sample of Queries Generating Matrices

Number of Queries	Number of Cases	Number of Queries	Number of Cases
1	15	1	257
1	57	1	265
1	59	1	321
1	64	1	521
1	149	1	528
1	149	1	616
1	150	1	3299
1	150	1	3712
1	158	1	4089
1	202	1	22380

20 jobs, average of 1857 database cases read
19 jobs, excluding 22380, average of 777 cases read

database cases meeting the selection criteria. The average number of cases displayed by this group is 265.

Like the ground database, the characterization of the matrix jobs was done by selecting a random sample of the matrix queries and running them. For the selected queries, an average of 1857 database cases were read. Removing the one sample which retrieved all the accidents between October 1, 1983 and May 17, 1988, the average becomes 777 cases. Table 6.11 includes the counts of database cases retrieved for each job run.

6.7.2.2. ARPS Performance

Like the ground database, the log data for aviation contained no information about the amount of time required to complete a query using the current ARPS program. A model for the current ARPS program was developed to provide a standard for comparison. The ARPS program for aviation is more complicated than the program for the ground database. The data for the ground database is stored in a single file (excluding narrative data) while the aviation data is stored in five files. Each aviation case has a basic data record, but only some cases have miscellaneous, impact, expanded personnel and 3W data. To provide a rough estimate of the performance of the current aviation program, only the time to access the basic file was included. This is an accurate estimate for approximately 40 percent of the queries in the evaluation set and is an under estimate for the remaining 60 percent.

The time derived is based on the wall clock time to do a query on an otherwise idle system. Thus the time includes both CPU and I/O times.

The assumptions for the model are:

- Cases to be searched are specified by the start and end dates included in the query specification.
- Database cases are uniformly distributed over time.
- No cost for formatting the output for the user is included in the time estimates. Checking the cost for formatting the non-matrix output revealed that the output formatting added less than 0.5 percent on average to the calculated times.

- Time for a case can be approximated by calculating the time to do only the basic record for that case. Time to read the miscellaneous, impact, expanded personnel and/or 3W data is not included.

  For each case the model does the following operations for each case:

1) Read the record.
2) Evaluate the selection criteria to determine if the record is selected or not.

To obtain numbers to use with this model, sample runs were made. Each sample run retrieved all the records from January 1, 1972 to the present and produced a matrix which included the total number of records retrieved. The record count plus the wall clock time for the retrieval was used to determine the number of records per second which could be read and processed. The time to read the aviation basic record was 2 minutes and 31 seconds. Using the same formula used for the ground database (see Section 6.7.1.2), estimated performance for the current ARPS program was derived. Table 6.12 is the performance for the queries which produce non-matrix output. Table 6.13 is the performance for the queries which produce matrix style output.

TABLE 6.12. Estimated Time to Complete Non-Matrix Queries Using Model of Current ARPS Program

No. of Queries	No. of Seconds	Cumulative Percent	No. of Queries	No. of Seconds	Cumulative Percent
205	9.41	56	2	34.14	84
33	9.43	65	2	37.65	85
2	9.48	66	1	41.56	85
1	9.51	66	1	42.71	86
1	9.54	66	1	43.45	86
1	9.56	66	1	44.41	86
1	9.66	67	1	47.06	86
1	9.69	67	1	50.58	87
1	9.74	67	1	52.92	87
1	9.78	67	1	52.93	87
2	9.82	68	1	52.96	87
1	9.84	68	1	52.98	88
1	9.87	69	2	53.11	88
3	9.9	69	1	53.16	89
1	10.05	70	1	55.64	89
1	10.21	70	6	56.43	90
1	10.22	70	7	56.46	92
1	10.58	70	1	56.88	93
1	10.97	71	1	60.19	93
2	14.74	71	1	60.35	93
2	15.33	72	1	65.83	93
1	15.49	72	3	66.42	94
1	16.09	72	2	67.27	95
22	18.8	78	1	71.92	95
3	18.83	79	1	81.18	95
1	18.84	80	2	88.28	96
1	22.75	80	2	88.31	96
2	24.69	80	2	88.44	97
1	24.71	81	1	97.61	97
1	24.74	81	1	97.82	98
5	25	82	1	106.96	98
1	25.11	83	1	107.25	98
1	25.88	83	1	116.63	98
1	31.84	83	4	144.93	99
2	31.97	84	1	163.45	100
1	32.15	84	1	163.84	100

Average time: 21 seconds

TABLE 6.13. Estimated Time to Complete Matrix Queries Using Model of Current ARPS Program

No. of Queries	No. of Seconds	Cumulative Percent	No. of Queries	No. of Seconds	Cumulative Percent
1	5.48	5	1	28.21	55
1	5.93	10	3	43.55	70
1	6.13	15	4	47.02	90
6	9.4	45	1	56.43	95
1	15.59	50	1	62.52	100

Average time for all 20 queries: 28 seconds
Average time for 19 queries: 27 seconds

6.7.2.3. Structure of the Database

The aviation database was analyzed to determine an appropriate relational structure. Review of the current structure indicated that it was a good first approximation to the correct organization for a relational database. Thus, the current aviation database structure was used for modelling. This does not imply that the current structure should be used when the database is actually implemented. The relations identified are:

- Basic data. This includes fields tagged as B, A, M and P in the aviation data dictionary.
- Miscellaneous data. This includes fields tagged as C in the data dictionary.
- Expanded personnel data. This includes fields tagged as I and E.
- Impact data. This is the fields tagged as D.
- 3 W data. This is the fields tagged as W.
- Narrative data. This is the textual description of the accident and is broken into 13 fields. Examples are Findings and Synopsis.

6.7.2.4. Expected Database Performance

The method for analyzing the aviation database is exactly the same as that used for the ground database. Appendices C.2.1, C.2.2 and C.2.3 contain detailed information about the fields used in the selection criteria and displayed to the user.

For the non-matrix queries, the average time was 20 seconds; Table 6.14 shows the distribution. For the matrix queries, the average time for a sample of 20 queries was 142 seconds. Removing the one matrix query which retrieved all of the accidents since October 1, 1983, the average time for the matrix queries is 61 seconds. Table 6.15 shows the distribution.

TABLE 6.14. Estimated Time to Complete Non-Matrix Queries Using the DBMS Model

No. of Queries	I/O Time No. of Seconds	Cumulative Percent	No. of Queries	I/O Time No. of Seconds	Cumulative Percent
176	3.75	48	1	82.5	96
76	7.5	69	2	101.25	96
36	11.25	79	1	105	97
13	15	82	1	123.75	97
7	18.75	84	1	127.5	97
15	22.5	88	1	131.25	98
5	33.75	90	1	135	98
9	37.5	92	1	172.5	98
2	45	93	1	191.25	98
1	52.5	93	1	221.25	99
3	56.25	94	1	225	99
2	63.75	94	1	303.75	99
3	67.5	95	1	457.5	99
1	75	95	1	555	100
1	78.75	96	1	607.5	100

Average time: 20 seconds

TABLE 6.15. Estimated Time to Complete Matrix Queries Using the DBMS Model

No. of Queries	I/O Time No. of Seconds	Cumulative Percent	No. of Queries	I/O Time No. of Seconds	Cumulative Percent
1	0.375	5	1	7.45	60
1	1.425	10	1	8.025	65
1	1.475	15	1	13.025	70
1	1.6	20	1	15.4	75
1	3.725	25	1	26.4	80
1	3.95	30	1	82.475	85
2	3.75	40	1	92.8	90
1	5.05	45	1	102.225	95
1	6.425	50	1	559.5	100
1	6.625	55			

Average time for all 20 queries: 142 seconds
Average time for 19 queries: 61 seconds

6.7.2.5. Analysis of Database Performance

Comparison of the ARPS performance and the projected database performance shows little difference in the time required to handle the non-matrix queries: 21 seconds for ARPS and 20 seconds for the database (see Tables 6.12 and 6.14). Comparison of Tables 6.13 and 6.15 shows a definite difference in the handling of matrix queries. ARPS processes the 19 queries in the random sample in 27 seconds. The projected database time is 61 seconds, or 2.3 times longer. For purposes of the comparison below, the non-matrix times will be judged to be equal. The non-matrix queries are 67 percent and matrix queries are 33 percent of the total aviation database queries. Applying the following calculation and considering the load caused by ARPS to be 1.0, the load caused by the projected database implementation is then 1.5 times the ARPS load.

%-non-matrix * speed-factor + %-matrix * speed-factor = DBMS load

.67 * 1.0 + .33 * 2.5 = 1.5

The method of prediction used in section 6.7.2.4 specified that the location of the database cases could be done entirely with keys into the database. This is true in the organization described in Section 6.7.2.3 only if the selection criteria can be evaluated using fields from only one of the relations listed. This is true in 93 percent of the queries in the evaluation set. To allow 98 percent of the queries to be done entirely with keys, one new relation is needed, the combination of basic and expanded personnel data. The remaining 2 percent of the queries would still take longer than predicted. Of these 0.5 percent use the miscellaneous data and the remainder use the 3W data.

Note that this would NOT be the method used to implement the database. This is for estimation only and serves to identify structure within the aviation database which must be more fully studied and characterized before the aviation data is placed in a relational DBMS. More careful characterization of these queries will identify the underlying structure.

The database design described in Section 6.7.2.4 specified no physical storage constraints. In particular, there is nothing known about the relationship between the location of the basic data and the expanded personnel, miscellaneous, impact or 3W data for an accident. This implies that queries that get data from more than one relation, require more time than queries which can be totally satisfied by one relation. By providing constraints on the physical location for the data for a case, we can reduce the time required. Recomputing the times for both the matrix and non-matrix to reflect the change in the physical storage results in the matrix jobs taking an average of 58 seconds and the non-matrix jobs taking an average of 9 seconds. Redoing the calculation above gives:

%-non-matrix	*	speed-factor	+	%-matrix	*	speed-factor	=	DBMS load
.67	*	.43	+	.33	*	2.1	=	.98

Like the ground database, there are a number of users who use the database heavily. Review of all sessions with more than 10 queries per session revealed 32 matrices which could use already selected data. This reduces the matrix load to 84 percent of the original and the DBMS load then becomes:

%-non-matrix	*	speed-factor	+	%-matrix	*	speed-factor	=	DBMS load
.67	*	.43	+	.33	*	2.1 * .84	=	.87

This is not as large a reduction as in the ground database.

The above DBMS model is based on I/O time to obtain the requested records. In the aggregate, this model allows as much CPU time as I/O time because when multiple queries are running simultaneously, one query is using the CPU while others are using the I/O subsystem. The projected DBMS load indicates that the DBMS-based system could use 1.2 times as much CPU as I/O and still take only 1.04 times as long as the current ARPS system.

The 591 queries were analyzed to see what the role of previously selected data might be. At the end of each query the user is asked what he wishes to do next and if appropriate whether he wishes to do further displays based on the data already selected. One hundred fifty-seven (157) queries (of a total 591) had an already selected set associated and the user indicated that the data should be reused. This means that 26 percent of the all queries could be satisfied using data already selected by the DBMS. This does not imply that the already selected data can necessarily be used directly. The current ARPS implementation filters these cases again using the selection criteria the user specifies. Thus the user is free to further restrict the selection.

Looking more closely at the set of queries which had an already selected set shows these 157 "reuse" responses were 48 percent of the queries where the "do further displays based on the data already selected" was appropriate. Reusing already selected data is appropriate if you are not setting up a batch job and if this is not the last query in your session. Thus 48 percent of the time the user could have used the already selected data subset. This clearly indicates that there is a pattern to the queries submitted by a single person and thus the new implementation of ARPS should use this knowledge to reduce the load on the DBMS.

6.7.3. FECA, Flying Hours and Exposure Databases

The three remaining databases generated only 8 percent of the queries in the evaluation set. No performance comparison was done. In each database, the data would logically be stored in a single relation and thus tuning the database by adding keys will be the method for gaining the necessary speed.

As with the other two databases, the work of generating statistical tables should be moved to another package which can more efficiently do this work.

6.7.4. Summary of Performance Evaluation

A database management system can do what ARPS is currently doing, but only if it is careful planned and implemented. The analysis of the DBMS based system proposed an organization for each database and showed that the organization did not correctly reflect all uses. Further analysis of the data should be done to identify the additional structure in the data so it can be reflected in the database organization.

There are two components to the ability to do this restructuring. One component is a detailed knowledge of the data. You must understand what is in the data and how it is used. The USASC staff has a good understanding of both the data and the user's requests. The other component of this ability is a knowledge of the structure and workings of the DBMS. This information can be learned. All vendors offer courses which provide this information. For a DBMS implementation to be successful, the USASC staff will need to learn and apply these skills.

Analysis of the query patterns revealed a substantial portion of queries are based on the data selected for the previous query. Thirty-five (35%) percent of all ground queries and 26 percent of all aviation could use data selected by the previous query. Looking at only the subset of queries where a subset is available indicates that 62 percent of the ground and 48 percent of the aviation queries could use previously selected data. This knowledge should be used to structure the user's work environment to make it easy to use previously selected data and thus reduce the load on the DBMS. To make the reuse of the same data available not only during a query session but spanning query sessions, the data selected should be saved on disk if the user requests.

The generation of statistical tables could more efficiently be done by a package specifically designed for this purpose. SAS on the IBM 4381 or a personal computer based package such as SAS for the PC, P-STAT, SPSS PC+, etc., should be used. To make a DBMS-based implementation of ARPS successful, it should include more than a single program, specifically it should use the

DBMS to select data and other programs (on the IBM 4381 or on personal computers) to analyze that data.

The largest benefit would come from moving this analysis work to personal computers. This can not be a requirement because the speed of communications for dial-in access makes the transfer of data to a remote PC much less viable. Also no limitation has been put on the types of device necessary to access the ASMIS system. Those users who do not have a PC or who are using dial-in lines can still use packages on the IBM 4381 to do this work.

To make this multi-program, multi-computer structure viable for the users, the transfer to these other programs must be easy. Realizing the potential in this portion of the solution will take an educational effort. To assess the impact of this change, the number of queries per user group was analyzed. Because the tools to do non-matrix displays may well be different from the tools used for the generation of statistical table, the two types of queries, matrix and non-matrix were analyzed separately. As can been seen in Table 6.16 for the ground database and Table 6.17 for the aviation database, approximately 85 percent of the queries for each of the databases come from only two or three organizations. Thus training is practical.

TABLE 6.16. Percent of Total Queries by User Group for the Most Frequent Users of the Ground Database

	USASC	FCHQ	EUHQ
Percent of queries	73	11	2
Percent of matrix queries	60	23	5
Percent of lines printed by non-matrix queries	47	14	0

TABLE 6.17. Percent of Total Queries by User Group for the Most Frequent Users of the Aviation Database

	USASC	DCAS
Percent of queries	70	17
Percent of matrix queries	64	20
Percent of lines printed by non-matrix queries	55	20

The current ARPS system is a relatively neutral environment for the user because the complexity of the query does not affect the time required to obtain the data. The only penalty that the user notices is the number of days of data to be searched. For a fixed date interval, the complexity of his query does not affect the time required. This is particularly true of the ground database where all the data for each case is read independent of what portion of the case the query requires. This is less true in the aviation database because the data for a case is in more than one file. Using a DBMS, there will be penalties for complex queries (look at the distribution of query times in Table 6.6 versus those in Table 6.8 for instance). The DBMS provides response to uncomplicated queries more quickly than the current ARPS, but it responds to complex queries more slowly than the current ARPS. The users can learn what these penalties are and simply adjust their query accordingly. Here again is a place for user training, support and guidance.

6.8. DISK SPACE UTILIZATION

The disk space requirements for the ASMIS database were reviewed. Table 6.18 shows the space requirements for the data in the current databases. Allowing for increase in space required when using a DBMS (2 to 3 times the actual data) results in the collective database (both on-line and off-line data) occupying approximately 30 to 45 percent of the currently available disk space. Adding another 2.5 gigabyte disk, results in the database occupying 23 to 34 percent of the available disk space. From the estimate of growth rate, the yearly increase in space required would be 2.3 to 3.4 percent of the 10 gigabyte configuration.

Review of this analysis by Jim Hayes indicated that approximately one third of the disk space is occupied by files necessary for computer operation, including MVS/SP, SAS, ARPS executables, etc. Efforts should be made to reduce this requirement, thus allowing more disk space for data storage and manipulation. If nothing can be done to reduce this space requirement, then the current on-line database occupies 15 percent of the currently available disk space (that is 15 percent of 5 gigabytes). Adding the off-line data and expanding this by 2 to 3 times as required by using a DBMS would result in 45 to 68 percent of the currently available disk space for data (an unacceptably high percentage). Adding another 2.5 gigabyte drive (bringing the system to 10 gigabytes with 7.5 gigabytes available for data storage) results in the data occupying 30 to 45 percent of the space with an annual growth rate of 3.0 to 4.5 percent. This is still a relatively high percentage and More disk space would be advisable. Addition of a second 2.5 gigabyte drive (for a total of 12.5 gigabytes with 10 gigabytes available for data storage) would allow 23 to 34 percent of disk space for the DBMS-based database with an average growth rate of 2.3 to 3.4 percent. The addition of this second drive may require an additional disk controller.

Establishing the DBMS-based database in an environment where 23 to 34 percent of the disk space is for the DBMS with a growth rate of 2.3 to 3.4 percent would provide 3 to 4 years of expansion capacity before any other solution would be necessary.

The current on-line database under ARPS occupies only 10% of the currently available disk space and yet programmers report that disk space is always tight. One explanation is that the IBM 4381 is being used for other things (this is not obvious in the performance analysis) and the disk space is occupied by those tasks. Alternative explanations are a large degree of duplication of data or a large quantity of unused files being maintained on disk.

TABLE 6.18. Disk Space Required for the Current ASMIS Databases

File Name	Number of Records Online	Number of Records Offline	Aver Rec Len	Size Megabytes Online	Size Megabytes Offline	Size Megabytes Total	Growth Rate Records	Growth Rate Megabytes
GROUND DATABASE								
Basic Info	143000	120000	642	91.81	77.04	168.85	20000	12.84
Narrative	660000	550000	252	166.32	138.60	304.92	92000	23.18
3W file	107000		242	25.89	0.00	25.89		
Exposure	27000		240	6.48	0.00	6.48		
Total				290.50	215.64	506.14		36.02
AVIATION DATABASE								
Basic Info	66000	31000	779	51.41	24.15	75.56	4800	3.74
Miscellaneous	36000	49000	333	11.99	16.32	28.30	2600	0.87
Personnel	52000		1420	73.84	0.00	73.84	3800	5.40
Impact	1500		360	0.54	0.00	0.54	100	0.04
Narrative	91000		806	73.35	0.00	73.35	6600	5.32
Dash-2	1000		80	0.08		0.08		
Cross Ref	193000		60	11.58		11.58	4800	0.29
Flying Hours	78000		80	6.24		6.24	114000	9.12
Total				229.03	40.47	269.49		24.76
FECA DATABASE								
monthly	60000	60000	520	31.20	31.20	62.40	20000	10.40
quarterly	112000	112000	420	47.04	47.04	94.08	38000	15.96
Total				78.24	78.24	156.48		26.36
DRUG and ALCOHOL DATABASE								
CODARS	286000	85000	457	130.70	38.84	169.55	60000	27.42
Urinalysis	128000	0	40	5.12	0.00	5.12		
DA3711-R	5300	0	742	3.93	0.00	3.93		
Civdaa	1	0	281	0.00	0.00	0.00		
Total				139.75	38.84	178.60		27.42
MISCELLANEOUS FILES								
UIC Trans	8100		95	0.77	0.00	0.77		
Statistics	143000		114	16.30	0.00	16.30		
Total				17.07	0.00	17.07		
TOTAL for all ASMIS databases				754.59	373.19	1127.79		114.57
Percent of available (7.5 gigabytes)				10.06	4.98	15.04		1.53

REFERENCES

Babcock, C. 1987. "Monsanto Weighs Oracle Although DB2's Faster." Computer World, November 23, 1987, pp. 6.

Bucken, M. June 1988. "DB2 is Faster, More Dominant?" Software Magazine, 8(8):32.

DATAPRO Research Corporation. March 1988. "Cullinet Software, Inc. IDMS/R - C/ICMS" DATAPRO, pp. SW25-251MJ-101 - SW25-251MJ-106.

Inmon, B. May 9, 1988. "Rating the Vendors: Whose DBMS Runs the Fastest?" Computer World, 22(19):81-85.

Levine, R. November 1988. "Optical Storage Comes of Age." DEC Professional, 7(11): pp. 49-59.

Schussel, G. 1986. "IBM's Database and Programmer Productivity Aids." In Proceedings of the 1986 National Database and Fourth Generation Language Symposium, pp. 58-79, Digital Consulting Associates, Inc.

Stoll, C. May 1988. "Stalking the Wiley Hacker." Communications of the ACM, 31(5), pp. 484-497.

APPENDIX A

GLOSSARY

APPENDIX A: GLOSSARY

ADAPCP	Alcohol and Drug Abuse Prevention and Control Program.
ANSI	American National Standard Institute.
ARPS	ASMIS Retrieval and Processing System.
ASMIS	Army Safety Management Information System.
Batch mode	A non-conversational method of obtaining information from a computer. The user poses a question and returns later to receive his response. Typically delay between query and response is measured in hours.
Change control	A method of managing and documenting changes. For the ASMIS application, this means keeping detailed records of what changes were made, when and why they were made. Computer programs exist which will facilitate this work.
CODARS	Client Oriented Drug and Alcohol Reporting System.
CPU	Central Processing Unit.
Database	A database is a collection of data which provides a complete description of a related group of entities. For the current ASMIS implementation, a database is the collection of files describing an accident type, e.g., ground, and providing auxiliary data needed to calculate accident rates, translate codes, etc.
Data independence	An application is independent of the data it uses if changes can be made to the data, its storage location, etc., with minimal impact on the application.
DBA	Database administrator. The person who manages a database. He analyzes the significance of proposed changes and supervises the implementations; authorizes users to access portion(s) of the database; makes backups of the database; and monitors the database for performance and performs the necessary tuning.
DDN	Defense Data Network.
DOIM	Office of Information Management.
FECA	Federal Employees Compensation Act.

Interactive mode	A conversational method of obtaining information from a computer. The user poses a question and the computer responds in a relatively short period of time. Response times vary form a few seconds to a few minutes.
Integrity	The problem of integrity is the problem of ensuring that the database contains only accurate data. Inconsistency between two entries which represent the same "fact" is an example of lack of integrity. For example, a lack of integrity occurs in the ground database when the sum of the individual injury costs (fields PINJCOST) don't equal the total injury cost for the accident (field INJCOST). This type of lack of integrity can be eliminated by removing the redundancy; i.e., by calculating total injury cost. Even if redundancy is eliminated, the database may still contain inaccurate data. An example is hours on duty = 500. This type of lack of integrity can be eliminated by validating the data as it is entered and/or modified.
IPR	In Progress Report.
MACOM	Major command.
Matrix output	In the context of the current implementation of ARPS, matrix output is a two-dimensional table which relates the values of one field to the values of a second field (ARPS also offers a three-dimensional matrix). In many statistical analysis packages, this is called a crosstabulation or crosstabs.
Nonmatrix output	In the content of the current implementation of ARPS, nonmatrix output is a columnar listing of the fields requested in the query specification. Each line of the output represents one database case.
PC	Any personal computer.
PROC	Specifications for the ARPS query processor. A PROC may specify a whole query or only parts of the query (e.g. specification of fields to be displayed and summed). These specifications are generated by hand and thus a limited number exist. Knowledge of the existence of these PROCs is very limited.

Query	The means of obtaining information from a database. A query consists of three parts: the selection criteria, the selected fields and subsequent processing to be done with the selected fields for the selected cases. Options for subsequent processing include report generation or the generation of a data subset. As used in conjunction with the ASMIS system, a query can be specified using the ARPS program or verbally. Verbal queries are converted to ARPS queries by a USASC staff member and the report is returned to the requestor, either verbally or in written form.
RAID	Office of Research, Analysis and Investigation.
Relation	A structural entity in a relational database. This is a collection of data fields related to a single concept. An example is the personal data in the ground database.
Report	One type of output from a query. A report displays the selected fields for each case matching the selection criteria. Display means print the exact value, translate the code into more easily understood text and print the text, or apply an algorithm to all selected cases and print the result (e.g., sum, average), etc.
Selected fields	The fields that are available for further processing, e.g., to be included in a report, made available in a data subset.
Selection Criteria	The user's statement which specifies how to select data from the database.
SMD	Office of Systems Management.
SNA	System Network Architecture.
USASC	US Army Safety Center.
USADAOA	USA Drug and Alcohol Operations Activity.
User	Anyone, either inside or outside the USASC, who needs information from the ASMIS system. If it is necessary, internal and external will be used to distinguish the two types of users.

APPENDIX B

CURRENT FORMS

* U.S. Government Printing Office 1981 - 355424

INSTRUCTIONS FOR DA FORM 285

GENERAL. The unit having the accident must investigate it and complete this report. Complete only starred items for nonfatal off-duty accidents not involving Army operations or materiel. For all other accidents, complete all items except those for safety staff or Safety Center use only. Type or print the report. Items may be continued on an attached sheet. Items not in the instructions are self-explaining.

1a. Enter the six-part unit identification code *(UIC)* of the unit having the accident.

1b. Enter the description of the unit. For example, enter HHC 2/34 Inf, 194 CAV, Yuma PG.

2. If unknown, estimate.

3. Dawn is between first light and official sunrise. Dusk is between official sunset and night.

4. "On Post" means the accident happened on property under Department of Defense control.

5. Enter facts needed to locate the accident scene. As needed, enter building number or direction and distance from closest land mark; enter street or highway name or number, enter city or military installation; enter state and country.

SECTION A – PERSONNEL INVOLVED

Complete this section for each person involved in the accident. "Involved" means a person who was injured or who caused or contributed to the accident. If more than one person was involved, use more forms and complete only this section on them. Witnesses and uninjured passengers are not considered involved. Be sure to also complete this section on each supervisor who caused or contributed to the accident. Give the supervisory error in Item 30. In case of damage to property with no personnel involved *(e.g., fire, natural disaster)*, report only Items 6, 7, and 8 for the property custodian or the hand-receipt holder.

7. Give official address for all Government personnel. Leave out for all others. Include the unit UIC if different from the UIC in Item 1.

8. Complete for all Government personnel. Leave out for all others.

9. Enter pay grade for all Government personnel including foreign national employees. For example, enter E6, O4, WG8, GS-12, C-5A. Leave out for all others.

10-13. Complete for Government personnel only. Leave out for all others.

14. "On Duty" means: *(a)* person was at duty station during duty hours; or *(b)* person was away from duty station during duty hours but on official business. Leave out for non-Government personnel.

15-16. Complete for Government personnel only. Leave out for all others.

18. Enter this person's activity or task. For example, enter firing rifle, lifting box, walking across street, driving truck.

19-21. Leave out if activity *(Item 18)* was not required for training. For example, exclude horseplay, chow run, stand down.

22. Pick the term below that best describes the overall mission of the activity or task in Item 18.

Administrative; office	Medical
Maintenance; repair, services	Physical training; recreation
Transportation; supply, disposal	Food Services
Production; construction	Other operation
Research; development, testing	Personal; domestic
Emergency services; law enforcement	Off duty

23. The following definitions apply:

b. Permanent total disability means person can never again do gainful work.

c. Permanent partial disability means person loses or can never again use a body part.

d. Lost workday case - days away from work means person misses one or more days of work.

e. Lost workday case - restricted work activity means person is temporarily unable to perform regular duties.

f. Nonfatal case without lost workday means person *(1)* was permanently transferred or terminated, *(2)* received treatment greater than first aid, *(3)* lost consciousness, or *(4)* had an occupational illness that did not result in fatality or lost workday.

g. First aid only means one time treatment of minor injuries.

24. Estimate the number of workdays this person will lose. Do not update this estimate unless this person dies.

25. Estimate the number of workdays this person cannot perform all regular duties after going back to work.

26. Describe this person's injury or occupational illness. For example, enter third-degree chemical burn, first-degree thermal burn, compound fracture, dermatitis, heatstroke, concussion.

27. For the injury or illness shown in item 26, give the body part involved. For example, enter left knee, lungs, right thumb, nose.

28. Pick from the list below the event that resulted in the injury or illness. Then give the thing that produced it. For example, enter struck against door; bodily reaction due to slip; overexertion due to lifting box; exposure to noise.

Struck against ...	Bodily reaction due to...
Struck by ...	Overexertion ...
Fell from elevation onto ...	Exposure to ...
Fell from same level onto ...	External contact with ...
Caught in/under/between ...	Ingested ...
Rubbed/abrated by ...	Inhaled ...

30. For each mistake this person made, pick one error from the list below. Use error in a sentence that includes the result of the error. For example, due to *improper attention*, SGT Jones did not yield the right of way to the other vehicle. PFC Smith made an *improper decision* to drive while under the influence of alcohol; Mr English *failed to follow procedures (SOP)* and began spot welding without his safety goggles in place. Due to *inadequate planning* by the company commander *(CPT Wright)*, there was no unit ice and snow removal program. As a result, PFC Carr broke his arm by falling on the icy steps.

Inadequate inspection	Failed to comply with general rules/principles
Improper attention	Improper simple physical action *(lift, hold, drop, hit, push, pull, sit, stand reach for, open, close, connect, disconnect, etc.)*
Failed to recognize	Improper complex physical action *(walk, run, crawl, climb, carry, jump, align, adjust, steer, brake, etc.)*
Misjudged clearance/speed/weight/size	Inadequate communication *(ask, answer, signal, inform, etc.)*
Misinterpreted	
Failed to anticipate	
Inadequate planning	
Improper decision	
Inadequate improvising/troubleshooting/problem solving	
Failed to follow procedures/orders/laws	

SECTION B – PROPERTY AND/OR MATERIEL INVOLVED

31a. List all property involved in the accident whether damaged or not. For example, enter Tank, M60A1. "Property Involved" means materiel which is damaged or whose use or misuse contributed to the accident.

31b. Give ownership for each item listed. For example, enter Army, Air Force, Army National Guard, contractor, or private.

31c. If accident involved Army operations, enter estimated total cost of damage. Total will include costs of parts and labor.

32. For each materiel failure or malfunction, pick one type from the list below. Use the type in a sentence to tell how the materiel failed. Include nomenclature of materiel as in item 31. For example, M60A1 fuel line connector *vibrated*, loose and sprayed fuel over engine causing fire. F1500M road grader fuel brake master cylinder rubber piston seal *decayed* and failed causing loss of fluid and brake failure.

Overheated/burned/melted	Pulled/stretched
Froze *(temperature)*	Twisted/torqued
Obstructed/pinched/clogged	Compressed/hit/punctured
Vibrated	Bent/warped
Rubbed/worn/frayed	Sheared/cut
Corroded/rusted/pitted	Decayed/decomposed
Overpressured/burst	Electric current action *(short, arc, surge, etc.)*

33. TM 38-750 requires a Category I EIR for materiel failures or malfunctions that cause or contribute to accidents.

SECTION C – ENVIRONMENTAL CONDITIONS INVOLVED

34. For each environmental condition, pick one type from the list below. Use the type in a sentence that describes its role in the accident. For example, driver's vision was restricted by fog; air breathed was *contaminated* by toxic fumes; heat exhaustion resulted from high *temperature*; person slipped and fell on *floor* made *slippery* by wax.

Illumination *(dark, glare, etc.)*	Radiation *(sunlight, x-ray, LASER, etc.)*
Precipitation *(rain, fog, ice, snow, etc.)*	Work surface *(slippery floor, cluttered walkway, steep rough road, etc.)*
Contaminants *(fumes, dust, chemicals, FOD, etc.)*	Air pressure *(explosion, decompression, altitude effects, etc.)*
Noise	Electricity *(lightning, arc, surge, short, shock, etc.)*
Temperature/humidity	
Wind/turbulence	
Vibration	
Acceleration/deceleration	

SECTION D – DESCRIPTION AND CORRECTIVE ACTION

35. Give the sequence of events that describes what happened leading up to and including the accident. In describing cause factors be sure to *(a)* name personnel making errors, *(b)* tell how involved personnel are related to materiel listed in Item 31, e.g., passenger in M151A2 or lighting immersion heater, and *(c)* tell how environmental conditions affected personnel or materiel. Continue on an attached sheet if necessary.

37. This item is to be completed by the commander or his representative.

38. Command review as locally required.

SAFETY STAFF USE ONLY

GENERAL. The safety staff will complete this section on all accidents. The safety staff will investigate all accidents requiring a DA Form 285-1 and will attach it to this report.

39. When change is checked, Items 1, 2, 6, and 8 must be completed plus any changes.

40. Enter MACOM of the unit shown in Item 1. For example, enter FORSCOM, TRADOC, USAREUR, NGB, or COE.

42. From the list below, select the type that best describes this accident. Types are listed in order of precedence to help pick one when more than one applies.

Army motor vehicle	Fire
Army combat vehicle	Chemical
Army operated vehicle	Explosive
Privately owned vehicle	Missile
Marine owing	Radiation
Marine underway	Nuclear
Marine not underway	Personnel injury – other
Other Army vehicle	Property damage – other

43. Describe the type of vehicle collision. For example, ran off road and overturned, head-on collision, sideswipe, or vehicle struck pedestrian.

UNITED STATES ARMY INVESTIGATION ACCIDENT REPORT

For use of this form, see AR 385-40; the proponent agency is [illegible]

REQUIREMENT CONTROL SYMBOL [illegible]

NOTE: SPACES BELOW, DEFINED BY HEAVY LINES ARE FOR "SAFETY CENTER USE ONLY"

1. UNIT IDENTIFICATION: a. UIC | b. DESCRIPTION

2. TIME AND DATE OF ACCIDENT: a. YEAR | b. MONTH | c. DAY | d. HOUR

3. TIME OF DAY (Check one): ☐ a. DAWN ☐ b. DAY ☐ c. DUSK ☐ d. NIGHT

4. LOCATION: ☐ a. ON POST ☐ b. OFF POST

5. EXACT LOCATION OF ACCIDENT

SECTION A — PERSONNEL INVOLVED

6. NAME (Last, First, MI)

7. ADDRESS [illegible]

8. SOCIAL SECURITY NUMBER

9. GRADE

10. AGE

11. SEX: ☐ a. MALE ☐ b. FEMALE

12. MOS OR CIVILIAN JOB SERIES

13. FLIGHT STATUS: ☐ a. YES ☐ b. NO

14. DUTY STATUS: ☐ a. ON DUTY ☐ b. OFF DUTY

15. NO. OF HOURS ON CONTINUOUS DUTY BEFORE ACCIDENT

16. NO. OF HOURS SLEEP IN LAST 24 HOURS [illegible]

17. CLASSIFICATION AT TIME OF ACCIDENT (Check appropriate box)

☐ a. ACTIVE ARMY ☐ e. OTHER US MILITARY
☐ b. ARMY CIVILIAN ☐ f. ROTC
☐ c. ARMY CONTRACTOR ☐ g. DEPENDENT
☐ d. NONAPPROPRIATED FUND

NATIONAL GUARD: ☐ h. TECH ☐ i. IDT ☐ j. AT ☐ k. FTTD ☐ l. FTM ☐ m. ADT

ARMY RESERVE: ☐ n. IDT ☐ o. AT ☐ p. ADT ☐ q. FTM

FOREIGN NATIONAL: ☐ r. DIRECT HIRE ☐ s. CONTRACT HIRE ☐ t. KATUSA

☐ u. OTHER (Specify)

18. THIS PERSON'S ACTIVITY/TASK AT TIME OF ACCIDENT

19. IF THIS PERSON'S ACTIVITY WAS NECESSARY PART OF TRAINING, GIVE TYPE: ☐ a. BASIC (School) ☐ b. ADVANCED (School) ☐ c. OJT (Unit) ☐ d. PROFICIENCY (Unit) ☐ e. OTHER (Specify)

20. WAS THIS PERSON'S ACTIVITY PART OF FIELD EXERCISE? ☐ a. YES ☐ b. NO

21. WAS THIS PERSON'S ACTIVITY PART OF TACTICAL TRAINING? ☐ a. YES ☐ b. NO

22. OPERATIONAL CATEGORY (Identify operational category that best describes the overall mission at time of accident.)

23. SEVERITY OF INJURY TO THIS PERSON (Check only one)

☐ a. FATAL ☐ b. PERMANENT TOTAL DISABILITY ☐ c. PERMANENT PARTIAL DISABILITY ☐ d. LOST WORKDAY — DAYS FROM WORK
☐ e. LOST WORKDAY CASE — RESTRICTED WORK ACTIVITY ☐ f. NONFATAL CASE WITHOUT LOST WORKDAYS
☐ g. FIRST AID ONLY ☐ h. NO INJURY ☐ i. MISSING AND PRESUMED DEAD

24. WORKDAYS LOST (number)

25. WORKDAYS RESTRICTED (number)

26. TYPE/NATURE OF INJURY/OCCUPATIONAL ILLNESS

27. BODY PART AFFECTED

28. CAUSE OF INJURY/OCCUPATIONAL ILLNESS

29. VEHICLE RESTRAINT SYSTEM: ☐ a. USED ☐ b. NOT AVAILABLE ☐ c. NOT APPLICABLE ☐ d. AVAILABLE BUT NOT USED

30. THIS PERSON'S ERRORS WHICH CAUSED OR CONTRIBUTED TO THE ACCIDENT (Describe each mistake and the results)

SECTION B — PROPERTY AND/OR MATERIAL INVOLVED

31. LIST ALL PROPERTY INVOLVED IN THE ACCIDENT WHETHER DAMAGED OR NOT. IF ACCIDENT INVOLVED ARMY OPERATIONS, SHOW COST OF ANY DAMAGE

NO	a. NAME OF ITEM (Complete nomenclature, i.e., name, type, model)		b. OWNERSHIP		c. AMOUNT OF DAMAGE
1					
2					
3					

32. MATERIAL FAILURES/MALFUNCTION(S) WHICH CAUSED OR CONTRIBUTED TO THE ACCIDENT (Tell what failed and how it failed)

33. CONTROL NUMBER FOR THE EIR COVERING EACH FAILURE/MALFUNCTION (Block 3 of SF 368)

SECTION C — ENVIRONMENTAL CONDITIONS INVOLVED

34. ENVIRONMENTAL CONDITION(S) WHICH CAUSED OR CONTRIBUTED TO THE ACCIDENT

SECTION D — DESCRIPTION AND CORRECTIVE ACTION

35. FULLY DESCRIBE THE ACCIDENT (When material is listed in item 31, tell how involved personnel are related to it.)

36. ACTION TAKEN, ANTICIPATED, OR RECOMMENDED TO CORRECT THE CAUSE(S) OF THIS ACCIDENT

37. SIGNATURE OF COMMAND REPRESENTATIVE

38. COMMAND REVIEW: a. | b. | c.

SAFETY STAFF USE ONLY

39. REPORT SUBMISSION: ☐ a. INITIAL ☐ b. CHANGE

40. MACOM

41. LOCAL REPORT NUMBER

42. ACCIDENT TYPE

43. TYPE OF VEHICLE COLLISION

44. SAFETY STAFF POINT OF CONTACT (Include printed name and phone)

45. SPECIAL REQUIREMENTS

46. DATE REPORT COMPLETED (Yr-Mo-Day)

DA FORM AUG 80 285 EDITION OF 1 JUN 71 IS OBSOLETE.

ACCIDENT INVESTIGATION REPORT						REQUIREMENT CONTROL SYMBOL CSGPA-147(R4)	
1. LOCAL REPORT NUMBER	2. TIME AND DATE OF ACCIDENT					3. UNIT UIC	4.
	a. YEAR	b. MONTH	c. DAY	d. HOUR			PAGE OF PAGES
5. TELL WHAT HAPPENED	6. TELL WHAT CAUSED/ALLOWED IT TO HAPPEN					7. TELL WHAT TO DO ABOUT IT	

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART I - STATEMENT OF REVIEWING OFFICIALS

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL
CSGPA - 1551

1. REVIEWING OFFICIALS COMMENTS

2. APPROVING AUTHORITY COMMENTS

a. SIGNATURE

3. DEPARTMENT OF ARMY REVIEW

a. SIGNATURE

4. CASE NUMBER				USASC USE ONLY	
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.		DELETE	1.
				ADD	2.
				CHANGE	3.

DA FORM 2397-R, MAR 83

REPLACES DA FORM 2397, AUG 80, WHICH IS OBSOLETE.

TECHNICAL REPORT OF U.S. ARMY AIRCRAFT ACCIDENT
PART II - SUMMARY

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA-1551

1. CLASSIFICATION	2. TYPE EVENTS	3. TIME OF DAY
☐ A ☐ B ☐ C ☐ D ☐ E	a. b. c.	1 ☐ DAWN 2 ☐ DAY 3 ☐ DUSK 4 ☐ NIGHT
1. ☐ ON POST ON AIRFIELD 2. ☐ ON POST NOT ON AIRFIELD 3. ☐ ON AIRFIELD OF ANOTHER SVC 4. ☐ ON CIVIL AIRFIELD 5. ☐ OFF POST NOT ON AIRFIELD	5. NEAREST MILITARY ESTABLISHMENT 6. TOTAL NUMBER OF AIRCRAFT INVOLVED	7. LOCATION a. GRID COORDINATES b. CITY, STATE, COUNTRY

8. a. MISSION, TYPE, DESIGN, SERIES	10a. ESTIMATED COSTS ☐ TOTAL LOSS
b. ORGN AIRCRAFT ASSIGNED *(UIC)*	ACFT DAMAGE COST $ OWNER
c. INSTAL AIRCRAFT ASSIGNED	REPAIR M/HRS $
9. ORGANIZATION/CHAIN OF COMMAND DEEMED MOST	OTHER DAMAGE MIL $ OWNER
RESPONSIBLE/CAPABLE OF TAKING CORRECTIVE ACTION	OTHER DAMAGE CIV $ OWNER
UNIT/UIC | UNIT/UIC | UNIT/UIC | UNIT/UIC | UNIT/UIC | MACOM/UIC	INJURY COST $
	TOTAL COST THIS ACFT $
	b. TOTAL COST MULTIPLE ACFT EVENT $

11. SURVIVABILITY	12. INFLIGHT ESCAPE	13. FIRE	14. POST CRASH ESCAPE DIFFICULTIES	15. FUEL	LB FUEL
1. ☐ SURVIVABLE	1. ☐ EJECTION	0. ☐ NONE		a. AT TAKE OFF	
2. ☐ PARTIALLY SURV	2. ☐ BAILOUT	1. ☐ INFLIGHT		b. AT TIME OF EMERG.	
3. ☐ NON SURVIVABLE	3. ☐ NOT ACCOMPL.	2. ☐ POST CRASH	1. ☐ YES	c. TERMINATION	
4. ☐ ACFT MISSING	4. ☐ NA	3. ☐ OTHER	2. ☐ NO		

16. FLAMMABLE FUEL SPILLAGE		17. CLEARANCE	
NONE	0 ☐	VFR	0 ☐
FUEL	1 ☐	IFR	1 ☐
ENGINE OIL	2 ☐	NONE	2 ☐
HYDRAULIC FLUID	3 ☐	FROM	TO
TRANSMISSION OIL	4 ☐		
CARGO	5 ☐		
UNDETERMINED	9 ☐	18. MISSION	
OTHER *(Specify)*	8 ☐		

19. INJURIES *(Number)*	FATAL A	DISABLING B - E	NONDISABLING F - G	MISSING, PRESUMED DEAD H	NOT INJURED J
a. OCCUPANTS MILITARY					
b. OCCUPANTS OTHER					
c. NON-OCCUPANTS MIL					
d. NON-OCCUPANTS OTHER					
e. TOTAL THIS ACFT					
f. MULTIPLE ACFT EVENT					

20. TERRAIN OF CRASH SITE *(More than one may apply)*

a. GEN CHARACTERISTICS	b. AT MISHAP SITE	c. SURFACE AT MISHAP SITE	d. OBSTACLES AT MISHAP SITE
14 ☐ MOUNTAIN 08 ☐ FLAT TERRAIN	12 ☐ LEVEL	01 ☐ PREPARED 04 ☐ ICE	17 ☐ STUMPS 05 ☐ TREES
13 ☐ DESERT	07 ☐ SLOPE	02 ☐ SOD 15 ☐ SNOW	10 ☐ BLDG 18 ☐ WIRES
11 ☐ ROLLING 09 ☐ WATER		03 ☐ SOGGY 16 ☐ WATER	06 ☐ ROCKS/BOULDERS 98 ☐ OTHER

21. FLIGHT DATA

	FLIGHT DURATION	PHASE OF OPERATIONS	ALTITUDE AGL	ALTITUDE MSL	AIRSPEED KIAS	HEADING *(Compass)*	AIRCRAFT WEIGHT	DENSITY ALTITUDE	OVERGROS YES	OVERGROS NO
a. PLANNED	HR TNS									
b. WHEN EMERGENCY OCCURRED	HR TNS									
c. ACCIDENT SEQUENCE TERMINATION	HR TNS									

22. ACCIDENT CAUSE FACTORS *(Enter a "D" or "S" in appropriate blocks to identify definite or suspected causes)*

a. PERSONNEL		PERSONNEL *(Continued)*	
(1) FLIGHT CREW: DUTY		(3) SUPERVISORY DUTY	
DUTY		DUTY	
DUTY		(8) OTHER DUTY	
(2) GROUND CREW: DUTY		b. MATERIAL FAILURE/MALFUNCTION	
DUTY		c. ENVIRONMENTAL	

23. SEQUENCE *(Enter a concise summary of accident sequence from onset of emergency through termination of flight. See DA Pam 385-95 for sample statement and restrictions on length of statement)*

24. CASE NUMBER a. DATE *(YYMMDD)*	b. TIME	c. ACFT SERIAL NO.	25. AVIATION SAFETY OFFICER *(Name, orgn and signature)*	26. OTHER ACFT SERIAL NUMBER

DA FORM 2397-1-R, MAR 83

REPLACES DA FORM 2397-1, AUG 80, WHICH IS OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT
PART III - FINDINGS AND RECOMMENDATIONS

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is OCSPER.

REQUIREMENT CONTROL SYMBOL CSGPA - 1551

1. FINDINGS AND RECOMMENDATIONS *(Attach additional sheet, if required)*

2. SUMMARY OF ACCIDENT CAUSES, SYSTEM INADEQUACIES AND RECOMMENDATIONS

		SYSTEM INADEQUACIES	REMEDIES		
a. PERSONNEL ERROR		1.	1.	2.	3.
DUTY CODE		2.	1.	2.	3.
TASK ERROR CODE		3.	1.	2.	3.
b. PERSONNEL ERROR		1.	1.	2.	3.
DUTY CODE		2.	1.	2.	3.
TASK ERROR CODE		3	1.	2.	3.
c. PERSONNEL ERROR		1.	1.	2.	3.
DUTY CODE		2.	1.	2.	3.
TASK ERROR CODE		3.	1.	2.	3.
d. MATERIAL FAILURE/MALFUNCTION		1.	1.	2.	3.
FAILURE CODE		2.	1.	2.	3.
		3.	1.	2.	3.
e. ENVIRONMENTAL		1.	1.	2.	3.
ENVIRONMENTAL CODE		2.	1.	2.	3.
		3.	1.	2.	3

3. CASE NUMBER				USASC USE ONLY	
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.		DELETE	1.
				ADD	2.
				CHANGE	3.

DA FORM 2397-2-R, MAR 83

REPLACES DA FORM 2397-2, JUL 74 AND DA FORM 2397-9, JUL 74, WHICH ARE OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART IV - NARRATIVE

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL
CSGPA - 1551

1. NARRATIVE ACCOUNT OF INVESTIGATION *(Use format shown in DA Pamphlet 385-95)*

2. CASE NUMBER				USASC USE ONLY	
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.		DELETE	1.
				ADD	2.
				CHANGE	3.

DA FORM 2397-3-R, MAR 83

REPLACES DA FORM 2397-3, JUL 74, WHICH IS OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART V - SUMMARY OF WITNESS INTERVIEW

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL CSGPA - 1551

1. NAME OF WITNESS *(Last, first, MI)*	2. OCCUPATION/TITLE	3. GRADE	4. SSN	5. AGE

6. ADDRESS *(Include ZIP Code) (If military, include organization)*	7. TELEPHONE NUMBER
	8. DATE OF INTERVIEW

9. AVIATION EXPERIENCE AND BACKGROUND	10. LOCATION	11. INTERVIEWER

12. WITHIN THE ARMY, THIS STATEMENT WILL BE USED SOLELY FOR ACCIDENT PREVENTION/SAFETY PURPOSES AND IT MAY NOT BE USED AS EVIDENCE OR TO OBTAIN EVIDENCE IN ANY JUDICIAL, ADMINISTRATIVE, OR DISCIPLINARY ACTION OR PROCEEDING; IN DETERMINING MISCONDUCT OR LINE-OF-DUTY STATUS OF PERSONNEL; BEFORE FLIGHT EVALUATION BOARDS IN DETERMINING LIABILITY IN CLAIMS FOR OR AGAINST THE GOVERNMENT; IN DETERMINING PECUNIARY LIABILITY (AR 385-40).

13. CASE NUMBER		
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.

DA FORM 2397-4-R, MAR 83 REPLACES DA FORM 2397 4, JUL 74, WHICH IS OBSOLETE

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART VI - WRECKAGE DISTRIBUTION

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL
CSGPA - 1551

1. GRID: SHOW MAJOR GOUGE MARKS, DISTRIBUTION OF WRECKAGE, OBSTACLES, DIRECTION OF NORTH, WIND DIRECTION, WIND VELOCITY, POSITION OF WITNESS, ETC. SUGGESTED SCALE: 1" EQUALS 40' ACTUAL SCALE: 1" EQUALS ________

2. CASE NUMBER			3. OTHER AIRCRAFT SERIAL NUMBER
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.	

DA FORM 2397-5-R, MAR 83 REPLACES DA FORM 2397-5, JUL 74, WHICH IS OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART VII - IN-FLIGHT OR TERRAIN IMPACT AND CRASH DAMAGE DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA - 1551

1. INFLIGHT COLLISION KINEMATICS AT INSTANT OF IMPACT

a. AIRSPEED AT IMPACT *(Knots)*

b. VERTICAL SPEED *(Feet per minute)* ________ . ☐UP ☐DOWN

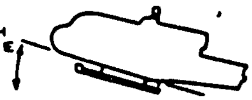
c. WIND VELOCITY AT IMPACT *(Knots)*

d. WIND DIRECTION AT IMPACT *(Degrees)*

e. FLIGHT PATH ANGLE *(Degrees)* ________ . ☐UP ☐DOWN

f. INFLIGHT ATTITUDE AT IMPACT

(1) PITCH ANGLE



(2) ROLL ANGLE

DEGREES ________ ☐UP ☐DOWN DEGREES ________ ☐L ☐R

g. OBSTACLE IDENTITY AND LOCATION

OBSTACLE	COLLISION HEIGHT ABOVE GROUND *(Feet)*
(1) ☐BIRDS	
(2) ☐AIRCRAFT	
(3) ☐WIRES/CABLES	
(4) ☐VEHICLES	
(5) ☐TREE	
(6) ☐OTHER	

h. OBSTACLE STRIKE SEQUENCE

(1) ☐PROP/ROTOR (2) ☐ROTOR MAST (3) ☐TAIL ROTOR (4) ☐TAIL BOOM (5) ☐WINDSCREEN/CANOPY (6) ☐LWR NOSE/GUN TURRET (7) ☐LANDING GEAR (8) ☐WING (9) ☐EMPENNAGE (10) ☐OTHER *(Specify)*

i. OBSTACLE CONSPICUITY *(Within avoidance distance from pilots position, the obstacle in its surroundings was obscured)*

(1) ☐COMPLETELY (2) ☐PARTIALLY (3) ☐NOT OBSCURED

j. WIRE OR CABLE DESCRIPTION

TYPE	DIA IN INCHES	NO. STRUCK
(1) POWER TRANSMISSION		
(2) TELEPHONE OR TV		
(3) BRACING *(Guy/Support)*		
(4) OTHER *(Specify)*		

(5) WIRE PROTECTION SYSTEM INSTALLED ☐YES ☐NO

2. TERRAIN COLLISION KINEMATICS AT INSTANT OF MAJOR IMPACT

a. GROUND SPEED AT IMPACT *(Knots)*

b. VERTICAL SPEED *(Feet per minute)* ________ . ☐UP ☐DOWN

c. FLIGHT PATH ANGLE *(Degrees)* ________ . ☐UP ☐DOWN

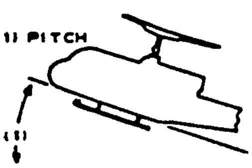
d. INDICATE BY CHECK MARKS WHICH TWO OF THE THREE PRECEEDING PARAMETERS *(a, b, c)* ARE THE MOST ACCURATE.

a. ☐ b. ☐ c. ☐

e. IMPACT ANGLE *(Degrees)*

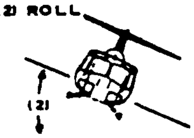
f. ATTITUDE AT MAJOR IMPACT

(1) PITCH



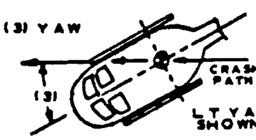
DEGREES ________ ☐UP ☐DOWN

(2) ROLL



DEGREES ________ ☐LEFT ☐RIGHT

(3) YAW



DEGREES ________ ☐LEFT ☐RIGHT

3. ROTATION AFTER MAJOR IMPACT

a. DID AIRCRAFT ROTATE ABOUT ANY AXIS AFTER THE ABOVE MAJOR IMPACT *(If yes, complete items b, c, and d)*

☐YES ☐NO ☐UNKNOWN

AIRCRAFT AXIS / ROTATIONS *(degrees)*	LEFT	RIGHT
b. ROLL		
c. YAW		

d. FORWARD NOSE OVER *(Degrees)*

4. IMPACT FORCES RELATIVE TO AIRCRAFT AXES *(G's)*

a. VERTICAL *(G's)* ________ . ☐UP ☐DOWN

b. LONGITUDINAL *(G's)* ________ . ☐FORE ☐AFT

c. LATERAL *(G's)* ________ . ☐LEFT ☐RIGHT

5. CASE NUMBER a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.	6. OTHER AIRCRAFT SERIAL NUMBER

USASC USE ONLY	
DELETE	1.
ADD	2.
CHANGE	3.

7. FUSELAGE INWARD DEFORMATION OR COLLAPSE AND INJURY RELATIONSHIP *(Check appropriate boxes)*									
FUSELAGE AREA	AMOUNT OR TYPE OF DEFORMATION OR COLLAPSE	SPECIFIC AREA OF DEFORMATION OR COLLAPSE				(5) FUSELAGE DEFORMATION PRODUCED/CONTRIBUTED TO INJURY			
		Cockpit (1)	Forward Cabin Area (2)	Mid Cabin Area (3)	Rear Cabin Area (4)	Cockpit	Forward Cabin Area	Mid Cabin Area	Rear Cabin Area
a. ROOF	UP TO 1 FOOT								
	MORE THAN 1 FOOT LESS THAN 3 FEET								
	MORE THAN 3 FEET								
b. LEFT SIDE	UP TO 1 FOOT								
	MORE THAN 1 FOOT								
c. RIGHT SIDE	UP TO 1 FOOT								
	MORE THAN 1 FOOT								
d. NOSE	UP TO 1 FOOT								
	MORE THAN 1 FOOT								
e. FLOOR	UP TO 1 FOOT								
	MORE THAN 1 FOOT								
f. FLOOR, *(Local deformation under seats)*	VERTICAL								
	SIDEWARD								
	FORWARD/REARWARD								

8. LARGE COMPONENT DISPLACEMENT *(Check appropriate boxes)*				
COMPONENT	DISPLACED (1)	TORN FREE (2)	PENETRATED/ENTERED	
			COCKPIT (3)	CABIN (4)
a. TRANSMISSION *(Forward or main)*				
b. TRANSMISSION *(Rear)*				
c. ROTOR BLADE *(Forward or main0*				
d. ROTOR BLADE *(Rear)*				
e. LANDING GEAR *(Specify location)*				
f. OTHER *(Specify)*				

9. POSTCRASH FLAMMABLE FLUID SPILLAGE

a. EQUIPPED WITH CRASHWORTHY FUEL SYSTEM ☐ YES ☐ NO

b. IF SO EQUIPPED DID BREAKAWAY VALVES SEPARATE ☐ YES ☐ NO

c. DID FLAMMABLE FUEL SPILLAGE OCCUR ☐ YES ☐ NO

d. WAS AIRCRAFT EQUIPPED WITH FIRE RESISTANT HYDRAULIC FLUID ☐ YES ☐ NO

e. AMOUNT AND TYPE FLUID SPILLED *(Check box)*

GALLONS	ENGINE FUEL	OIL	HYDRAULIC FLUID
0 – 1			
1 – 2			
2 – 10			
10 – 20			
20+			

f. SPILLAGE SOURCE

PART	PART NAME, TITLE, NOMENCLATURE	MANUFACTURERS NO.	NSN
(1) CELL/TANK/RESERVOIR			
(2) FILTER			
(3) FITTING			
(4) FLUID LINE			
(5) VALVE			
(6) BREAKAWAY VALVE			
(7) OTHER *(Specify)*			

10. REMARKS

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART VIII - MAINTENANCE AND MATERIEL DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA - 1551

1. AIRCRAFT HISTORY

a. ACCEPTANCE DATE *(YYMMDD)*		f. LAST INSPECTION	
b. AIRCRAFT ASSIGNED *(YYMMDD)*		DATE *(YYMMDD)*	
c. HOURS SINCE NEW		TYPE	
d. LAST MAJOR REPAIR FACILITY		g. HRS SINCE LAST INSPECTION	
		h. ORGN COMPLETING LAST INSP *(UIC)*	
e. HOURS SINCE LAST MAJOR REPAIR			

2. CAUSATIVE ROLE

	D	S	N
a. MATERIEL			
b. MAINTENANCE			
c. DESIGN			
d. MANUFACTURE			

3. FAILED OR MALFUNCTIONED MATERIEL

DID PART NUMBER MATCH THAT LISTED IN TM ☐ YES ☐ NO ☐ UNKNOWN

IDENTIFICATION	MAJ COMPONENT	PART	IDENTIFICATION	MAJ COMP	PART
a. NOMENCLATURE			(4) HRS SINCE NEW		
b. TYPE DESIGN, SERIES			(5) HRS SINCE LAST INSTALLED		
c. PART NUMBER			(6) DATE LAST INSTALLED *(YYMMDD)*		
d. NSN					
e. MFG CODE					
f. SERIAL NUMBER			(7) LAST OVERHAUL FACILITY		
g. TM DATA			(8) LAST SPECIAL INSP *(Type)*		
(1) TM NUMBER					
(2) DATE *(YYMMDD)*			(9) HOURS SINCE LAST SPECIAL INSP		
(3) FUNCTIONAL GP					
(4) FIGURE NUMBER			(10) DATE OF LAST SPECIAL INSP *(YYMMDD)*		
(5) ITEM NUMBER					
h. TAMMS DATA			i. TYPE/MODE OF FAILURE/ MALFUNCTION		
(1) NO. OF OVERHAULS					
(2) DATE OF LAST OVERHAUL *(YYMMDD)*			j. CAUSE OF FAILURE/ MALFUNCTION		
(3) HRS SINCE OVERHAUL			k. QDR/EIR NUMBER		

4. WARNING SYSTEM AND INDICATION OF FAILURE/MALFUNCTION

a. STATUS OF AIRCRAFT WARNING SYSTEM FOR THIS PART

b. INDICATIONS OF FAILURE/MALFUNCTION *(Enter from left to right in sequence they occurred)*

GENERAL		AIRCRAFT SYSTEM		
(1)	(2)	(3)	(4)	(5)

c. OP INDICATIONS AT TIME OF FAILURE/MALFUNCTION

(1) TORQUE		(6) ENGINE OIL TEMP	
(2) RPM (N^1)		(7) ENG OIL PRESSURE	
(3) RPM (N^2)		(8) EPR	
(4) RPM (r^2)		(9) FUEL FLOW	
(5) ENGINE EGT			

d. OTHER COMPONENT INDICATIONS

(1) TEMP	(3) RPM
(2) PRESSURE	(4) OTHER *(Specify)*

5. POL

a. POST ACCIDENT LAB RESULTS			
(1) TYPE			
(2) SOURCE			
(3) LOCATION OF LAB			
(4) DATE *(YYMMDD)*			
(5) FILTER CONDITION *(Specify)*			
(6) LAB RESULTS			
b. PRE-ACCIDENT LAB RESULTS			
(1) SPECIFY LAB			
(2) DATE LAST SAMPLE			
(3) LAB RESULTS			

6. TEARDOWN ANALYSIS

a. ORGN PERFORMING	b. SHIPPING INFORMATION	
c. MAINT REQ NO.	(1) SHIPPED *YYMMDD)*	(2) MODE
d. USASC CONTROL NO.	(3) GBL/BOL	(4) TCN NO.

7. REMARKS *(Use additional sheet if required)*

8. CASE NO.			9. OTHER AIRCRAFT SERIAL NUMBER
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.	

USASC USE ONLY	
DELETE	1.
ADD	2.
CHANGE	3.

DA 2397-7-R, MAR 83

REPLACES DA FORM 2397-7, JUL 74, AND DA FORM 2397-7A, JUL 74, WHICH ARE OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART IX - PERSONAL DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA - 1551

ROLE OF THIS INDIVIDUAL

COMMITTED ERRORS THAT CAUSED/CONTRIBUTED TO ACCIDENT (1) ☐ DEFINITELY (2) ☐ SUSPECTED (3) ☐ NO (4) ☐ UNKNOWN	b. AT CONTROLS WHEN ACCIDENT OCCURRED (1) ☐ YES (2) ☐ NO	c. DUTY STATUS (1) ☐ ON DUTY (2) ☐ OFF DUTY

2. BACKGROUND DATA

a. DATE LAST LEAVE ENDED *(YYMMDD)*		j. HOURS WORKED LAST 72 HOURS	
b. DAYS DURATION LAST LEAVE		k. DUTY HOURS REMAINING THIS DAY AFTER ACCIDENT OCCURRED	
c. HOURS SLEPT LAST 24 HOURS			
d. HOURS SLEPT LAST 48 HOURS		l. HEIGHT *(Inches)*	
e. HOURS SLEPT LAST 72 HOURS		m. WEIGHT *(Pounds)*	
f. HOURS AWAKE PRIOR TO ACCIDENT		n. AGE	
g. HOURS DURATION LAST SLEEP PERIOD		o. HOURS FLOWN LAST 24 HOURS	
h. HOURS WORKED LAST 24 HOURS		p. HOURS FLOWN LAST 48 HOURS	
i. HOURS WORKED LAST 48 HOURS		q. HOURS FLOWN LAST 72 HOURS	

3. RATED CREWMEMBER DATA

a. FW RATED *(YYMMDD)*		o. MTDS AIRCRAFT FLOWN LAST 60 DAYS ASP/IP	(1)
b. RW RATED *(YYMMDD)*			(2)
c. LAST PHYSICAL *(YYMMDD)*			(3)
d. WAIVERS ☐ YES ☐ NO		p. MTDS AIRCRAFT QUALIFIED/CURRENT IN	(1)
			(2)
			(3)
e. FAC 1 ☐ 2 ☐ 3 ☐ *(YYMMDD)*		q. ATM TASK NUMBER ASSOCIATED WITH INITIAL INDICATION OF EMERGENCY	
f. ARL 1 ☐ 2 ☐ 3 ☐ *(YYMMDD)*		(1) LAST PERFORMED *(YYMMDD)*	
g. −10 EXAMINATION *(YYMMDD)*		(2) NUMBER OF ITERATIONS	
h. ANNUAL WRIT *(YYMMDD)*		r. ATM TASK NUMBER INVOLVED IN RESPONSE TO EMERGENCY	
i. INSTRUMENT RENEWAL *(YYMMDD)*			
j. SDZN RENEWAL *(YYMMDD)*		(1) LAST PERFORMED *(YYMMDD)*	
k. MOST RECENT EVALUATION FLIGHT IN MISHAP MTDS AIRCRAFT *(YYMMDD)*		(2) NUMBER OF ITERATIONS	
		s. POST-ACCIDENT FLIGHT *(YYMMDD)*	
		RESULT	
l. NVG QUALIFIED ☐ YES ☐ NO		t. POST-ACCIDENT MEDICAL EXAMINATION/AUTOPSY *(YYMMDD)*	
m. IP ☐ SIP ☐ IFE ☐ MTP ☐ VT ☐		REQUIRED LAB TESTS ACCOMPLISHED ☐ YES ☐ NO	
n. PRIMARY AIRCRAFT MTDS		u. LOW PRESSURE/HIGH ALTITUDE CHAMBER ☐ YES ☐ NO	
		v. EJECTION SYSTEM QUAL ☐ YES ☐ NO	

4. FLYING EXPERIENCE

TYPE EXPERIENCE AND TIME	FIXED WING		ROTARY WING		TOTAL	WEATHER	MISHAP AIRCRAFT	
	SINGL ENG	MULTI ENG	SINGL ENG	MULTI ENG		INST	DESIGN	SERIES
a. INSTRUCTOR PILOT								
b. PILOT								
c. COPILOT								
d. CIVILIAN PILOT								
e. TOTAL TIME								
f. COMBAT TIME								
g. FLT SIMUL/SYNTH TRAINER								
h. TOTAL TIME LAST 30 DAYS								
i. TOTAL TIME LAST 60 DAYS								

j. MONTHLY FLIGHT HOURS PAST 12 MONTHS

(1) DATE *(YYMM)*												THIS MO.
(2) HOURS												

5. CASE NUMBER			6. OTHER AIRCRAFT SERIAL NUMBER	USASC USE ONLY	
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.		DELETE	1.
				ADD	2.
				CHANGE	3.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART X - INJURY/OCCUPATIONAL ILLNESS DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is OCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA - 1551

DEGREE OF INJURY

a. ☐ FATAL
b. ☐ PERMANENT TOTAL DISABILITY
c. ☐ PERMANENT PARTIAL DISABILITY
d. ☐ LOST WORKDAY CASE *(Days away from work)*
e. ☐ LOST WORKDAYS *(Days of restricted work activity)*
f. ☐ NONFATAL WITHOUT LOST WORKDAYS
g. ☐ FIRST AID ONLY
h. ☐ MISSING & PRESUMED DEAD

2. NUMBER OF LOST WORKDAYS

a. DAYS AWAY FROM WORK	b. DAYS HOSPITALIZED	c. DAYS RESTRICTED ACTIVITY

3. UNCONSCIOUS		4. AMNESIA		
a. HRS	b. MIN	a. ☐ RETROGRADE: HRS MIN	b. ☐ ANTEGRADE: HRS MIN	c. ☐ NONE

5. INJURIES

INJURY SEQ NO. a.	INJURIES: BODY REGION b.	ASPECT c.	BODY RGN QUALIFIER d.	INJURY TYPE RESULT e.	SEVERITY f.	FOR USASC USE (cost) g.	MECHANISM: ACTION h.	QUALIFIER i.	CAUSE FACTORS: SUBJECT j.	ACTION k.	QUALIFIER l.	FOR USASC USE (Weighted cost) m.

6. REMARKS *(Use additional sheet if required)*

7. AUTOPSY PERFORMED

a. ☐ YES b. ☐ NO c. ☐ PROTOCOL ATTACHED d. ☐ WILL BE FORWARDED

8. CAUSE OF DEATH	9. DUTY STATUS a. ☐ ON DUTY b. ☐ OFF DUTY

10. NAME	11. SSN	12. GRADE	13. SEX	14. DUTY	15. SVC	16. UIC

17. CASE NUMBER: a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.	18. OTHER AIRCRAFT SERIAL NUMBER

USASC USE ONLY	
DELETE	1.
ADD	2.
CHANGE	3.

DA FORM 2397-9-R, MAR 83

REPLACES DA FORM 2397-11, JUL 74, WHICH IS OBSOLETE.

7. MAINTENANCE AND SUPPORT PERSONNEL DATA		
a. PMOS		e. MOS VERIFICATION
(1) DATE AWARDED *(YYMMDD)*	(2) SOURCE ☐ OJT ☐ AIT ☐ CIVIL EXP ☐ UNK	(1) SQT ☐ GO ☐ NO GO (2) DEFINE TASK PERFORMANCE ☐ CORRECT ☐ INCORRECT ☐ NA (3) PERCENT GO ON SCOREABLE UNITS ______ % (4) OVERALL PERCENTILE ______ % (5) SQT WAIVERED ☐ YES ☐ NO
b. SMOS		
(1) DATE AWARDED *(YYMMDD)*	(2) SOURCE ☐ OJT ☐ AIT ☐ CIVIL EXP ☐ UNK	f. CIVILIAN JOB SERIES OR TITLE
c. DMOS		
d. DEFICIENT TASK NO.		(1) TASK RELATED TO JOB DESCRIPTION ☐ YES ☐ NO (2) PERFORMANCE STANDARDS FOR TASK ☐ YES ☐ NO
(1) DMOS RELATED ☐ YES ☐ NO (2) TASK INTERRUPTED OR DELAYED ☐ YES ☐ NO		

8. LABORATORY TESTS

TYPE TEST	SPECIMEN TESTED	RESULTS	NAME OF DRUG	USASC CODE BLOCK
a. CARBON MONOXIDE				
b. ALCOHOL				
c. DRUG SCREEN				
d. OTHER				

9. HISTORY OF DESEASES/DEFECTS

DIAGNOSIS	METHOD OF DISCOVERY				WAIVERS		USASC CODE BLOCK
	ANL PHY	SICK CALL	AUT-OPSY	OTHER	AUTH	DATE *(YYMMDD)*	

10. REMARKS

11. NAME *(Last, first MI)*	12. SSN	13. GRADE	14. SEX	15. DUTY	16. SVC	17. UIC

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART XI - PERSONNEL PROTECTIVE/ESCAPE/SURVIVAL/RESCUE DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENTS CONTROL SYMBOL CSGPA - 1551

1. DID THIS INDIVIDUAL SUSTAIN AN INJURY OR OCCUPATIONAL ILLNESS BECAUSE OF ACCIDENT ☐ YES ☐ NO

NOTE: If "YES" box is checked, insure a DA Form 2397-9-R is completed for this individual)

PERSONNEL PROTECTIVE/RESTRAINT/SURVIVAL EQUIPMENT

ITEM	TYPE (1)	REQUIRED (2)	NEEDED (3)	AVAILABLE (4)	USED (5)	PRODUCED, ALLOWED INJURY (6)	PREVENTED INJURY (7)	REDUCED INJURY (8)	FUNCTIONED AS DESIGNED (9)	INFORMATION CODES (10)			
a. HELMUT													
b. VISOR													
c. GLASSES													
d. FLIGHT SUIT													
e. FLIGHT GLOVES													
f. FLIGHT JACKET													
g. BOOTS													
h. OTHER CLOTHING													
i. LAP BELT													
j. SHOULDER HARNESS													
k. GUNNER HARNESS													
l. INERTIA REEL													
m. SEAT/LITTER													
n. SURVIVAL EQUIP													
o.													
p.													

3. PERSONNEL EVACUATION/ESCAPE	INFORMATION CODES				
a. METHOD OF ESCAPE					
b. LOCATION IN AIRCRAFT					
c. EXIT ATTEMPTED					
d. EXIT USED					
e. AIRCRAFT ATTITUDE DURING ESCAPE					
f. COCKPIT/CABIN CONDITIONS					
g. ESCAPE DIFFICULTIES					

4. LAPSED TIME FOR RESCUE

	DATE		HOUR OF DAY		LAPSED TIME	
	MM	DD	HR	MIN	HR	MIN
a. NOTIFICATION OF RESCUE PERSONNEL						
b. INDIV PHYSICALLY REACHED						
c. INDIV ACTUALLY ABOARD RESCUE VEH						
d. RESCUE COMPLETED/ABANDONED						

5. DISTANCE FROM ACCIDENT TO ACTUAL RESCUE VEHICLE AT TIME OF ACCIDENT

a. TO AIRCRAFT IN NAUTICAL MILES

b. TO GROUND VEHICLE IN STATUTE MILES

6. PERSONNEL SURVIVAL/RESCUE	INFORMATION CODES								
a. SURVIVAL PROBLEMS ENCOUNTERED									
b. MEANS USED TO LOCATE INDIVIDUAL									
c. RESCUE EQUIPMENT USED									
d. FACTORS THAT HELPED RESCUE									
e. FACTORS COMPLICATING RESCUE									
f. INDIVIDUAL PHYSICAL CONDITION									

g. VEHICLES ACTUALLY PERFORMING EVAC *(Specify)*

h. OTHER VEHICLE ASSISTING IN RESCUE *(Specify)*

7. REMARKS *(Use additional sheet, if required)*

8. NAME *(Last, first, MI)*	9. SSN	10. GRADE	11. SEX	12. DUTY	13. SVC	14. UIC

15. CASE NUMBER			16. OTHER AIRCRAFT SERIAL NUMBER	USASC USE ONLY	
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.		DELETE	1.
				ADD	2.
				CHANGE	3.

DA FORM 2397-10-R, MAR 83

REPLACES DA FORMS 2397-10, 2397-12, 2397-13, 2397-13A, 2397-14, AND 2397-14A, JUL 74, WHICH ARE OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART XII - WEATHER DATA

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL
CSGPA - 1551

1. ROLE OF WEATHER			
a. DEFINITE D	b. SUSPECTED S	c. NONE W	d. UNDETERMINED Z

2. GENERAL DATA AT TIME OF OCCURRENCE		
a. TEMPERATURE *(degrees Cent.)*	b. ALTIMETER SETTING *(HG)*	c. ALTIMETER READING *(Feet)*

3. SKY CONDITION	ACCIDENT SEQUENCE: INITIAL INDIC OF EMERG	AT EMERGENCY	DURING DESCENT	ACCIDENT OR TERMATION
a. ○CLEAR				
b. ⦶SCATTERED (feet)				
c. ◐BROKEN (feet)				
d. ⊕OVERCAST (feet)				
e. –X PARTIAL OBSCURATION				
f. X OBSCURATION				
x. UNKNOWN				
4. HORIZON				
a. VISIBLE				
b. PARTIALLY OBSCURED				
c. OBSCURED				
5. VISIBILITY *(Naut. miles)*				
6. OBSTRUCTION TO VISION				
a. NATURAL				
01. DUST				
02. FOG				
03. GROUND FOG				
04. HAZE				
05. ICE FOG				
06. SMOKE				
07. BLOWING DUST				
08. BLOWING SAND				
09. BLOWING SNOW				
00. NONE				
98. OTHER *(Specify)*				
b. INDUCED *(Rotorwash, etc.)*				
01. BLOWING SNOW				
02. BLOWING SAND				
03. BLOWING DUST				
04. BLOWING SPRAY				
00. NONE				
98. OTHER *(Specify)*				

7. WINDS	
a. ALOFT *(At enroute altitude)*	
(1) DIRECTION *(Degrees Mag.)*	(2) VELOCITY *(Kt)*
b. SURFACE WINDS	
(1) LANDING DIR. *(Degrees Mag.)*	(2) SURFACE WIND DIR. AND VARIANCE *(Degrees Mag.)*
(3) SURFACE WIND VELOCITY AND GUST SPREAD *(Kt)*	

8. AIRCRAFT ICING — NONE 0 ☐ YES 1 ☐

	ICING SEVERITY: TRACE (1)	LIGHT (2)	MODERATE (3)	SEVERE (4)
01. MAIN MOTOR BLADES				
02. WINGS				
03. PROPELLERS				
04. CONTROL SURFACES				
05. ROTOR HEAD				
06. TAIL ROTOR				
07. FUSELAGE				
08. PITOT STATIC SYSTEM				
09. CARBURETOR				
10. ENGINE AIR INLET				
11. FUEL VENTS				
12. ANTENNA				
13. WINDSCREEN				
98. OTHER *(Specify)*				

9. SIGNIFICANT WEATHER *(A maximum of three may be selected)*	ACCIDENT SEQUENCE: INITIAL INDIC OF EMERG	AT EMERGENCY	DURING DESCENT	ACCIDENT OR TERMINATION
01. HAIL				
03. SLEET				
05. ICE CRYSTALS				
06. DRIZZLE				
07. RAIN				
09. SNOW				
12. LIGHTNING				
13. THUNDER STORM				
14. FREEZING DRIZZLE				
15. FREEZING RAIN				
16. GUSTY WINDS				
97. UNKNOWN				
00. NONE				
98. OTHER *(Specify)*				

10. TURBULENCE — NONE 0 ☐ YES 1 ☐ *(If "YES" enter below "C" for continuous, "I" for intermittent, and "O" for occasional)*

a. LIGHT				
b. MODERATE				
c. SEVERE				
d. EXTREME				

11. FORECAST — CORRECT C ☐ INCORRECT I ☐ UNKNOWN U ☐

12. REMARKS *(Use additional sheet if required)*

13. CASE NUMBER		
a. DATE *(YYMMDD)*	b. TIME	c. AIRCRAFT SERIAL NO.

USASC USE ONLY	
DELETE	1.
ADD	2.
CHANGE	3.

DA FORM 2397-11-R, MAR 83

REPLACES DA FORMS 2397-16 AND 2397-17, JUL 74, WHICH ARE OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

PART XIII - FIRE DATA *(To be completed for all events involving fire)*

For use of this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL CSGPA - 1551

1. FIRE STARTED *(Check D — Definite S — Suspected)*	D	S
a. INFLIGHT		
b. UPON IMPACT *(Less than 1 minute)*		
c. UPON IMPACT *(More than 1 minute)*		
d. DURING REFUELING		
y. OTHER *(Specify)*		
z. UNDETERMINED ☐		

2. INDICATIONS OF FIRE *(More than 1 May apply, enter 1, 2 or 3 to show sequence)*

a. ☐ FIRE WARNING SYSTEM d. ☐ SMELL
b. ☐ OTHER INSTRUMENTS e. ☐ EXPLOSION *(Sound)*
c. ☐ SIGHT f. ☐ EXTERNAL COMMO
y. ☐ OTHER *(Specify)*

3. INITIAL AND PRINCIPAL LOCATION OF FIRE *(Enter 1 to indicate initial location, 2 to indicate principal location)*	D	S
01. ENGINE SECTION		
02. TRANSMISSION SECTION		
03. COCKPIT		
04. TAIL ASSEMBLY		
05. PASSENGER SECTION		
06. OXYGEN SYSTEM		
07. BAGGAGE COMPARTMENT		
08. EXTERNAL STORES		
09. FLARE POD		
10. ROCKET POD		
11. AMMUNITION STORES		
12. AVIONIC SECTION		
13. APU		
14. WHEEL WELL		
15. WHEEL BRAKE		
16. TAILPIPE		
17. INSTRUMENT PANEL		
18. BATTERY COMPARTMENT		
19. JUNCTION BOX		
20. HEATER COMPARTMENT		
21. FUEL CELL		
22. WING		
23. GUN TURRET		
24. TAIL BOOM		
25. CARGO SECTION		
26. TIRES		
98. OTHER *(Specify)*		
99. UNDETERMINED ☐		

4. IGNITION SOURCE	D	S
a. EXHAUST FLAMES		
b. SPARKS, FRICTION, e.g., SKIDDING		
c. ELECTRICAL SPARKS		
d. HOT SURFACES, e.g., EXHAUST DUCTS		
e. AIRCRAFT SUBSYSTEM		
f. AIRCRAFT OCCUPANT, e.g., LIGHTED CIGAR		
g. EXTERNAL OF AIRCRAFT, e.g., GRASS FIRE		
h. CARGO		
i. EXPLOSIVES		

4. IGNITION SOURCE *(Continued)*	D	S
j. SHORT CIRCUIT		
k. LIGHTNING		
l. STATIC ELECTRICITY		
m. OTHER *(Specify)*		
n. UNDETERMINED ☐		

5. COMBUSTIBLE MATERIAL	D	S
a. MAIN FUEL		
b. AUXILIARY FUEL		
c. HYDRAULIC FLUID		
d. ENGINE OIL		
e. TRANSMISSION OIL		
f. ELECTRICAL INSULATION		
g. ACOUSTICAL MATERIALS		
h. METAL *(Specify)*		
i. EXPLOSIVES		
j. UPHOLSTERY MATERIALS		
k. CARGO		
m. EXTERNAL MATERIAL *(Specify)*		
y. OTHER *(Specify)*		
z. UNDETERMINED ☐		

6. FIRE EXTINGUISHING SYSTEM	a. GND	b. AIRCRAFT INST	b. AIRCRAFT PORT
(1) NO EFFECT WHEN DISCHARGED			
(2) ACTIVATED, BUT DID NOT DISCHARGE			
(3) REDUCED FIRE			
(4) EXTINGUISHED FIRE			
(5) NOT ACTIVATED AND NOT NEAR FIRE			
(6) NOT ACTIVATED, BUT NEAR FIRE			
(7) NOT INSTALLED			

7. FIRE/SMOKE DETECTION SYSTEM	YES 1	NO 2	UN-DETM. 9
a. SYSTEM INSTALLED			
b. WARNING SYSTEM OPERATED PROPERLY			
c. SENSORS WITHIN RANGE OF			

8. EFFECT OF EMER SHUTOFF PROCEDURE *(Enter D, S, or Unk)*	ENG	FUEL	ELECT
a. EXTINGUISHED FLAME			
b. REDUCED FIRE			
c. NO EFFECTS			
d. NOT ACCOMPLISHED			
e. USED FAULTY PROCEDURE			

9. GENERAL DATA

a. EST OF AIRCRAFT FIRE DAMAGE *(Excl of impact damage)*
(1) ☐ 0-25% (2) ☐ 26-50% (3) ☐ 51-75% (4) ☐ 76-100%

b. FIRE DIMENSION: TO CLEAR FIRE, AIRCRAFT OCCUPANTS HAD TO MOVE *(Feet)*:

c. TOXICITY: WAS THERE EVIDENCE OF TOXIC PRODUCTS.
☐ 1 YES ☐ 0 NO *(If yes, name, co, etc.)*

d. DISTANCE TO NEAREST AVAIL MIL FIREFIGHTING EQUIPMENT (1) AIRMILES *(NM)* (2) ROAD MILES *(SM)*

e. IS AIRCRAFT EQUIPPED WITH CRASH RESISTANT
(1) FUEL CELLS ☐ 1 YES ☐ 0 NO (2) FUEL LINES ☐ 1 YES ☐ 0 NO

10. REMARKS *(Use separate sheet of paper)*

11. CASE NUMBER a. DATE *(YYMMDD)*	11. CASE NUMBER b. TIME	11. CASE NUMBER c. AIRCRAFT SERIAL NO.	12. OTHER AIRCRAFT SERIAL NUMBER	USASC USE ONLY	
				DELETE	1.
				ADD	2.
				CHANGE	3.

DA FORM 2397-12-R, MAR 83   REPLACES DA FORM 2397-15, JUL 74, WHICH IS OBSOLETE.

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT
INDEX A

or (this form, see AR 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER.

REQUIREMENT CONTROL SYMBOL
CSGPA - 1551

:SION, TYPE, DESIGN AND SERIES	2. CASE NUMBER

	INFORMATION	ENCL	NOT APPLIC.	SEE RE-MARKS
	COPY OF ORDERS APPOINTING INVESTIGATING BOARD			
	WEATHER REPORTS			
	CERTIFICATE OF DAMAGE/ECOD			
	DIAGRAMS AND/OR PHOTOGRAPHS			
	COPY OF EQUIPMT IMPROVEMENT REPT *(DA Form 2407)* QUALITY DEFICIENCY REPT *(SF 368)*			
	SPECIAL TECHNICAL REPORTS AND LABORATORY ANALYSIS			
	WEIGHT AND BALANCE *(DD Form 365F)*			
	COPY OF DIRECTIVES, REGULATIONS, ETC.			
	AUTOPSY REPORT *(DD Form 1322)*			
)	FLIGHT PLAN			
1	COPY OF ARMY AVIATORS FLIGHT RECORD *(DA Form 2408-12)*			
2	COPY OF AIRCRAFT INSPECTION AND MAINTENANCE RECORD *(DA Form 2408-13)*			
3	COPY OF UNCORRECTED FAULT RECORD *(DA Form 2408-14)*			
4	COPY OF EQUIP MODIFICATION RECORD *(DA Form 2408-5)*			
5	OTHER *(Specify)*			
16	OTHER *(Specify)*			
17	OTHER *(Specify)*			
18	OTHER *(Specify)*			

4. REMARKS

TECHNICAL REPORT OF U. S. ARMY AIRCRAFT ACCIDENT

INDEX B

For use of this form, see –R 385-40 and DA Pamphlet 385-95; the proponent agency is DCSPER

REQUIREMENT CONTROL SYMBOL CSGPA - 1551

1. MISSION, TYPE, DESIGN AND SERIES	2. CASE NUMBER

3. TAB	TITLE	DA FORM NUMBER	ENCL	NOT APPL	SEE RE-MARKS
a.	STATEMENT OF REVIEWING OFFICIALS	2397-R			
b.	SUMMARY OF MISHAP	2397-1-R			
c.	FINDINGS AND RECOMMENDATIONS	2397-2-R			
d.	NARRATIVE OF MISHAP	2397-3-R			
e.	WITNESS STATEMENTS	2397-4-R			
f.	WRECKAGE DISTRIBUTION DATA	2397-5-R			
g.	IN-FLIGHT OR TERRAIN IMPACT AND CRASH DAMAGE DATA	2397-6-R			
h.	MAINTENANCE AND MATERIEL DATA	2397-7-R			
i.	PERSONAL DATA	2397-8-R			
j.	INJURY/OCCUPATIONAL ILLNESS DATA	2397-9-R			
k.	PERSONAL PROTECTION/ESCAPE/SURVIVAL/RESCUE DATA	2397-10-R			
l.	WEATHER	2397-11-R			
m.	FIRE DATA	2397-12-R			

4. REMARKS

5. BOARD MEMBERS

	SSN	GRADE	BR	RATING	ADDRESS AND TEL NO.
a. PRESIDENT *(Name and Signature)*					
b. RECORDER *(Name and Signature)*					
c. FLIGHT SURGEON *(Name and Signature)*					
d. INSTRUCTOR PILOT *(Name and Signature)*					
e. MAINT OFFICER *(Name and Signature)*					
f. OTHER *(Name and Signature)*					

DA FORM 2397-14-R, MAR 83 REPLACES DA FORM 2397-19, JUL 74, WHICH IS OBSOLETE.

APPENDIX C

FIELDS USED IN ARGUMENTS AND FOR DISPLAY